

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2006, 12:45:45 ; Search time 96 Seconds
(without alignments)
2274.701 Million cell updates/sec

Title: US-10-804-785-2
Perfect score: 2744
Sequence: 1 QSACTLQSETHPPLTWQKCS.....TVCASGTTCQVLPYYSQCL 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2006, 12:46:45 ; Search time 21 Seconds
(without alignments)
1956.658 Million cell updates/sec

Title: US-10-804-785-2
Perfect score: 2744
Sequence: 1 QSACTLQSETHPPLTWQKCS.....TVCASGTTCQVLPYYSQCL 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2006, 12:46:45 ; Search time 21 Seconds
(without alignments)
1956.658 Million cell updates/sec

Title: US-10-804-785-2
Perfect score: 2744
Sequence: 1 QSACTLQSETHPPLTWQKCS.....TVCASGTTCQVLPYYSQCL 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:
1: /cgn2_6/ptodata/1/iaa/5_COMB.pep:
2: /cgn2_6/ptodata/1/iaa/6_COMB.pep:
3: /cgn2_6/ptodata/1/iaa/H_COMB.pep:
4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:
5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep:
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2006, 12:51:20 ; Search time 15 Seconds
(without alignments)
1533.568 Million cell updates/sec

Title: US-10-804-785-2
Perfect score: 2744
Sequence: 1 QSACTLQSETHPPLTWQKCS.....TVCASGTTCQVLPYYSQCL 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 235405 seqs, 46284737 residues

Total number of hits satisfying chosen parameters: 235405

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_New:
1: /SIDS5/ptodata/1/pubpaa/US08_NEW_PUB.pep1:
2: /SIDS5/ptodata/1/pubpaa/US06_NEW_PUB.pep:
3: /SIDS5/ptodata/1/pubpaa/US07_NEW_PUB.pep:
4: /SIDS5/ptodata/1/pubpaa/US08_NEW_PUB.pep:
5: /SIDS5/ptodata/1/pubpaa/PCT_NEW_PUB.pep:
6: /SIDS5/ptodata/1/pubpaa/US09_NEW_PUB.pep:
7: /SIDS5/ptodata/1/pubpaa/US09_NEW_PUB.pep1:
8: /SIDS5/ptodata/1/pubpaa/US10_NEW_PUB.pep:
9: /SIDS5/ptodata/1/pubpaa/US10_NEW_PUB.pep1:
10: /SIDS5/ptodata/1/pubpaa/US11_NEW_PUB.pep:
11: /SIDS5/ptodata/1/pubpaa/US11_NEW_PUB.pep1:
12: /SIDS5/ptodata/1/pubpaa/US60_NEW_PUB.pep:
*

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2006, 12:46:35 ; Search time 41 Seconds
(without alignments)
1166.335 Million cell updates/sec

Title: US-10-804-785-2
Perfect score: 2744
Sequence: 1 QSACTLQSETHPPLTWQKCS.....TVCASGTTCAVLPYYSQCL 497

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:
1: pir1:
2: pir2:
3: pir3:
4: pir4:
*

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2006, 12:45:59 ; Search time 230 Seconds
(without alignments)
1524.555 Million cell updates/sec

Title: US-10-804-785-2
Perfect score: 2744

Sequence: 1 QSACTLQSETHPPLTWQKCS.....TVCASGTTCQVLPYYSQCL 497
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 2166443 seqs, 705528306 residues
Total number of hits satisfying chosen parameters: 2166443
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : UniProt_05.80:*

RESULT 8

US-09-548-938A-10

; Sequence 10, Application US/09548938A
; Patent No. 6573086
; GENERAL INFORMATION:
; APPLICANT: EMALFARB, MARK AARON
; APPLICANT: BURLINGAME, RICHARD PAUL
; APPLICANT: OLSON, PHILIP TERRY
; APPLICANT: SINITSYN, ARKADY PANTELEIMONOVICK
; APPLICANT: PARRICHE, MARTINE
; APPLICANT: BOUSSON, JEAN CHRISTOPHE
; APPLICANT: PYNNONEN, CHRISTINE MARIE
; APPLICANT: PUNT, PETER JAN
; APPLICANT: VAN-ZEIJL, CORNELIA MARIA JOHANNA
; TITLE OF INVENTION: TRANSFORMATION SYSTEM IN THE FIELD OF FILAMENTOUS FUNGI
; FILE REFERENCE: 3123-4001
; CURRENT APPLICATION NUMBER: US/09/548,938A
; CURRENT FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Chrysosporium lucknowense
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (249)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (365)
; OTHER INFORMATION: Variable amino acid

US-09-548-938A-10

Query Match 61.8%; Score 1695; DB 2; Length 526;
Best Local Similarity 60.4%; Pred. No. 4e-124;
Matches 311; Conservative 68; Mismatches 112; Indels 24; Gaps 10;

Qy	1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL 60
	: : : : : : : ::
Db	18 QNACTLTAENHPSLTWSKCTSGSCTSVQGSITIDANWRWTHRTDSATNCYEGNKWDTSY 77
Qy	61 CPDNETCAKNCCLDGAAYASTYGVTSGNSLSIGFVTQSA-QKNVGARLYLMASDTTYQE 119
	: : : : :: : : :
Db	78 CSDGPSCASKCCIDGADYSSTYGITSGNSLNKFVTKQYSTNIGSRTYLMESDTKYQM 137
Qy	120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
	: :
Db	138 FQLLGNEFTFDVDVSNLGCGLNGALYFVSMADGGMSKYSGNKAGAKYGTGYCDSQCPRD 197
Qy	180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
	: : : : : : :: :
Db	198 LKFINGEANVENWQSSTDANAGTGKYGSCCSEMDVWEANNMAAFTPCHIPCXVIGQSRCE 257
Qy	240 GDGCGGTYSDNRYGGCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTFETSGA 299
	: : : : : : :
Db	258 GDSCGGTYSTDYAGICDPDGCFNSYRQGNKTFYKGK--MTVDTTKITVVTFQFLNSA 315
Qy	300 -----INRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGG-SSFSDKGGLTQFK 353
	: : : : : : : :
Db	316 GELSEIKRFYVQNGKVIPNSESTIPGVEGNSITQDWCDRQKAAGDVTDXQDKGGMVQMG 375
Qy	354 KATSGGMVLVMSIWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
	: : : : : : : : : : ::
Db	376 KALAGPMVLVMSIWDDHAVNMLWLDSTWPI-DGAGKPGAERGACPTTSGVPAEVEAEAPN 434

Db 79 CTDAKSCAQNCVVDGADYTSTYGITNGDSLKFKVKGQYSTNVGSRTYLMGEDKYQT 138
 Qy 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKPTNTAGAKYGTGYCDSQCPRD 179
 Db 139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMADGGLSRYPGNKAGAKYGTGYCDAQCPRD 198
 Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
 Db 199 IKFINGEANIEGTGSTNDPNAGAGRGTCCSEMDIWEANNMATAFTPHTPCTIIGQSRCE 258
 Qy 240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTFETS-- 297
 Db 259 GDSCGGTYSNERYAGVCDPDGCFNSYRQGNKTFYGKG--MTVDTTKITVVTFQLKDAN 316
 Qy 298 ---GAINRYVQNGVTQQPNAELGSYSGNELNDDYCTAEEAEFGG-SSFSDKGGLTQFK 353
 Db 317 GDLGEIKRFYVQDGKIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG 376
 Qy 354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
 Db 377 KALAGPMVLVMSIWDDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN 435
 Qy 414 AKVTFNSNIKFGPIGST-----GNPSGGNPPGGNPPGTTTRPATTGSSPGPTQS 464
 Db 436 SNVVFSNIRFGPIGSTVAGLPAGNGGNNGNPP---PPTTTSSAPATTASAGPKAG 492
 Qy 465 HYQCQGGIGYSGPTVCASGTTCQVLNPYYSQCL 497
 Db 493 RWQQCQGGIGFTGPTQCEEPYTCTKLNDWYSQCL 525

Title: US-10-804-785-2

RESULT 15

US-10-481-179-2

; Sequence 2, Application US/10481179
 ; Publication No. US20040197890A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Novozymes A/S
 ; APPLICANT: Lange, Lene
 ; APPLICANT: Wu, Wenping
 ; APPLICANT: Aubert, Dominique
 ; APPLICANT: Landvik, Sara
 ; APPLICANT: Schnorr, Kirk
 ; APPLICANT: Clausen, Ib
 ; TITLE OF INVENTION: Polypeptides having cellobiohydrolase I activity and
 ; TITLE OF INVENTION: polynucleotides encoding same
 ; FILE REFERENCE: 10129.204-WO
 ; CURRENT APPLICATION NUMBER: US/10/481,179
 ; CURRENT FILING DATE: 2003-12-17
 ; NUMBER OF SEQ ID NOS: 67
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 2
 ; LENGTH: 526
 ; TYPE: PRT
 ; ORGANISM: Acremonium thermophilum
 US-10-481-179-2

Query Match 65.6%; Score 1799; DB 4; Length 526;
 Best Local Similarity 64.5%; Pred. No. 2.8e-128;
 Matches 330; Conservative 57; Mismatches 107; Indels 18; Gaps 7;
 Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL 60
 Db 18 QQACTLTAENHPTLSWSKCTSGGSCTS VGSVTIDANWRWTHQVSSSTNCYTGNEDTSI 77

Qy	61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA-QKNVGARLYLMASDTTYQE	119
	: : : : :	
Db	78 CTDGASCAAACCLDGADYSPTYGITTSGNALSLQFVTQGPYSTNIGSRTYLMASDTKYQM	137
Qy	120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRD	179
	: :	
Db	138 FTLLGNEFTFDVDTGLGCGLNGALYFVSMDEDGGLSKYSGNKAGAKYGTGYCDSQCPRD	197
Qy	180 LKFINGQANVEGWEPPSSNNANTGIGGHGSCCSEMDIWEANSISEALTYPCTTVGQEICE	239
	: : : :	
Db	198 LKFINGEANNVGWTPSSNDKNAGLGNYGSCCSEMDVWEANSISAAYTYPCTTIGQTRCE	257
Qy	240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLLTVVTQFETSGA	299
	: : :	
Db	258 GDDCGGTYSTDRYAGECDPDGCDNSYRMGNTTFYKGK--MTVDTSKKFTVVTQFLTDSS	315
Qy	300 -----INRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSS-FSDKGGLTQFK	353
	: : : :	
Db	316 GNLSEIKRFYVQNGVVIPNSNSNIAGVSGNSITQAFCDAQKTAFGDTNVFDQKGGLAQM	375
Qy	354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN	413
	: :	
Db	376 KALAQPMLVMSLWDDHAVNMLWLDSTYPTN-AAGKPGAARGTCPTTSGVPADVESQAPN	434
Qy	414 AKVTFSNIKFGPIGST--GNPSGGNPPGGNPPGTTTRPATTGSSP-----GPTQSH	465
	: :	
Db	435 SKVIYSNIRFGPIGSTVSGLPGGGSNPGGSSSTTTTRPATSTTSSASSGPTGGTAAH	494
Qy	466 YGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL	497
	: :	
Db	495 WGQCGGIGWTGPTVCASPYTCQKLNDWYYQCL	526

Title: US-10-804-785-2

RESULT 4

S38794

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - imperfect fungus (*Humicola grisea*)

N; Alternate names: beta-glucan cellobiohydrolase; exoglucanase

C; Species: *Humicola grisea* var. *thermoidea*

C; Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C; Accession: S38794; S08240; A45869

R; Radford, A.

submitted to the EMBL Data Library, June 1991

A; Reference number: S38794

A; Accession: S38794

A; Molecule type: DNA

A; Residues: 1-525 <RAD>

A; Cross-references: UNIPROT:P15828; UNIPARC:UPI000012BE0F; EMBL:X17258; NID:g2760; PIDN:CAA35159.1; PID:g2761

A; Note: this is a revision to the sequence from reference S08240

R; de Oliveira Azevedo, M.; Radford, A.

Nucleic Acids Res. 18, 668, 1990

A; Title: Sequence of cbh-1 gene of *Humicola grisea* var. *thermoidea*.

A; Reference number: S08240; MUID:90175006; PMID:2308855

A; Accession: S08240

A; Molecule type: DNA

A; Residues: 1-299, 'H', 301-525 <DEO>

A; Cross-references: UNIPARC:UPI00001729F6; EMBL:X17258

A; Note: the authors translated the codon CAG for residue 87 as His

A; Note: this sequence has been revised in reference S38794

R; Azevedo, M.; de, O.; Felipe, M.S.S.; Astolfi-Filho, S.; Radford, A.

J. Gen. Microbiol. 136, 2569-2576, 1990

A; Title: Cloning, sequencing and homologies of the cbh-1 (exoglucanase) gene of *Humicola grisea* var. *thermoidea*.

A; Reference number: A45869; MUID:91178527; PMID:2127803

A; Accession: A45869

A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-20, 'R', 22-34, 'K', 36-86, 'H', 88-141, 'V', 143-157, 'Y', 159-237, 'QQH', 241-244, 'I', 246-299, 'H', 301-525 <AZE>
A;Cross-references: UNIPARC:UPI00001729F7; GB:M64588; GB:X17258
A;Note: this sequence has been revised. See entry S08240
C;Genetics:
A;Gene: cbh-1
A;Introns: 138/1
C;Superfamily: cellulose 1,4-beta-celllobiosidase I; fungal cellulose-binding domain homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;494-525/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 60.2%; Score 1652; DB 1; Length 525;
Best Local Similarity 57.3%; Pred. No. 1.8e-91;
Matches 294; Conservative 76; Mismatches 121; Indels 22; Gaps 7;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL 60
| ||::| :| || |::| ||::|| | : :||:||||| : ||||| || | :::
Db 19 QQACSLTTERHPSLSWNKCTAGGQCQTQVASITLDSNWRWTHQVSGSTNCYTGNKWDTSI 78

Qy 61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQK-NVGARLYLMASDTTYQE 119
| | ::||:|||:||| | ||||:|||:|||:|||: |||: ||||: ||| : |||
Db 79 CTDAKSCAQNCVGDGADYTSTYGITTNGDSLKFKVTKQHSTNVGSRTYLMGEDKYQT 138

Qy 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCP RD 179
| ||||||:||| : ||||||| |||||:|||:||| | ||||||| ||||:|||
Db 139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMADGGLSRYPGNKAGAKYGTGYCDAQCP RD 198

Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
:|||||:|||:||| | :| :| ||| :| :|||||:|||:| ||| :||| |||
Db 199 IKFINGEANIEGWTGSTNDPNAGAGRYGTCCSEMDIWEANNMATAFTPHPCTIIGQSRCE 258

Qy 240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTFETS-- 297
| | |||||: || | |||||: || | || :||| | || :|||:|||
Db 259 GDSCGGTYSNERYAGVCDPDGCFDFNSYRQGNKTFYGKG--MTVDTTKITVVTFQLKDAN 316

Qy 298 ---GAINRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGG-SSFSDKGGLTQFK 353
| | :|||:||| : :| :| :||| :| :||| :| :|||:|
Db 317 GDLGEIKRFYVQDGKIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG 376

Qy 354 KATSGGMVLVMSIWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
| | :| |||||:||| :| :|||:||| :| :| :|||:|||:|||:
Db 377 KALAGPMVLVMSIWDDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN 435

Qy 414 AKVTFNSNIKFGPIGST-----GNPSGGNPPGGNPPGTTTRPATTGSSPGPTQS 464
:| |||||:|||:||| | :||| | |||:|||:||| :| |||
Db 436 SNVVFSNIRFGPIGSTVAGLPAGNGGNNGNPP---PPTTTSSAPTTTASAGPKAG 492

Qy 465 HYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL 497
:| |||||:||| | ||| :|||:
Db 493 RWQQCGGIGFTGPTQCEEPYICTKLNDWYSQCL 525

Title: US-10-804-785-2

RESULT 6
S42093
cellulose 1,4-beta-celllobiosidase (EC 3.2.1.91) - Neurospora crassa
C;Species: Neurospora crassa
C;Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S42093
R;Taleb, F.; Radford, A.
submitted to the EMBL Data Library, February 1994
A;Description: Cloning sequencing and homologies of the CBH-1 (exocellulobiohydrolase) gene of Neurospora crassa.
A;Reference number: S42093

A;Accession: S42093
A;Molecule type: DNA
A;Residues: 1-516 <TAL>
A;Cross-references: UNIPROT:P38676; UNIPARC:UPI000011D714; EMBL:X77778; NID:g456657;
PIDN:CAA54815.1; PID:g456658

C;Genetics:

A;Introns: 227/3

C;Superfamily: cellulose 1,4-beta-celllobiosidase I; fungal cellulose-binding domain homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;485-516/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 56.9%; Score 1561; DB 2; Length 516;
Best Local Similarity 57.5%; Pred. No. 4.7e-86;
Matches 294; Conservative 62; Mismatches 129; Indels 26; Gaps 10;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL 60
| | || :: || || || : || | | : || : || || || || : || | || | | : ||

Db 18 QQAGTLTAKRHPSLTWQKCTRGGCPTLNT-TMVL DANWRWTHATSGSTKCYTGKQATL 76

Qy 61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNGARLYLMASDTTYQEF 120
|| | : || | || | | || : || | | : || | | || | | || | | | | | | |

Db 77 CPDGKSCAACNALDGAADTYGTGTYGITGSGWSLTLQFVTD---NVGARAYLMADDTQYQML 132

Qy 121 TLLGNEFSFDVDVSQQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRDL 180
|| | | || : | : || | | | | | : || | | | | | | | | | | | | | | | |

Db 133 ELLNQELWFDVDMNSIPCGLNGALYLSAMDADGGMRKYPTNKAGAKYATGYCDAQCPRDL 192

Qy 181 KFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEG 240
| : || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 193 KYINGIANVEGWTPSTNDAN-GIGDHGSCCSEMDIWEANKVSTAFTPHTCCTIEQHMCEG 251

Qy 241 DCGGGTYSNDNRYGGCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTQFETSGA- 299
| | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 252 DSCGGTYSDDDRYGVLCDADGCFNSYRMGNTTFYGEKG--TVDTSSKFTVVTQFIKDSAG 309

Qy 300 ---INRYYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGG-SSFSDKGGLTQFKK 354
| : | | | | | : | : | | | | | | | | | | | | | | | | | | | | | | | | |

Db 310 DLAEIKAFYVQNGKVIENSQSNDGVSGNSITQSFCKSQKTAFGDIDDFNKKGLKQMGK 369

Qy 355 ATSGGMVLVMSLWDDYYANMLWLDSYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNA 414
| : | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 370 ALAQAMVLVMSIWDDHAANMLWLDSYP---VPKVPGAYRGSGPTTSGVPAEV DANAPNS 426

Qy 415 KVTFSNIKFGPI-----GSTGNPSGNPPGGNPPGTTTRPATTGSSP-GPTQSHY 466
| | | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 427 KVAFSNIKFGHLGISPFSGGSSGTPP-SNPSSSASPSTSSTAKPSSTSTASNPSGTGAAHW 485

Qy 467 GQCGGIGYSGPTVCASGTTQVLPYYSQCL 497
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 486 AQCGGIGFSGPTTCPEPYTCAKDHDYISQCV 516

Title: US-10-804-785-2

RESULT 13

JE0313

exoglucanase (EC 3.2.-.-) - imperfect fungus (*Humicola grisea*)

C;Species: *Humicola grisea*

C;Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 09-Jul-2004

C;Accession: JE0313

R;Takashima, S.; Iikura, H.; Nakamura, A.; Hidaka, M.; Masaki, H.; Uozumi, T.

J. Biochem. 124, 717-725, 1998

A;Title: Isolation of the gene and characterization of the enzymatic properties of a major exoglucanase of *Humicola grisea* without a cellulose-binding domain.

A;Reference number: JE0313; MUID:98429588; PMID:9756616

A;Accession: JE0313

A;Status: preliminary

RP NUCLEOTIDE SEQUENCE.
 RA Furniss C.S.M., Williamson G., Kroon P.A.;
 RT "The substrate specificity and susceptibility to wheat inhibitor
 RT proteins of *Penicillium funiculosum* xylanases from a commercial enzyme
 RT preparation.";
 RL J. Sci. Food Agric. 85:574-582(2005).
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellobiose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 DR EMBL; AJ312295; CAC85737.1; -; Genomic_DNA.
 DR HSSP; Q09431; 1GPI.
 DR GO; GO:0005576; C:extracellular region; IEA.
 DR GO; GO:0016162; F:cellulose 1,4-beta-cellulobiosidase activity; IEA.
 DR GO; GO:0030248; F:cellulose binding; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR GO; GO:0030245; P:cellulose catabolism; IEA.
 DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
 DR GO; GO:0045493; P:xylan catabolism; IEA.
 DR InterPro; IPR000254; CBD_fun.
 DR InterPro; IPR001722; Glyco_hydro_7.
 DR Pfam; PF00734; CBM_1; 1.
 DR Pfam; PF00840; Glyco_hydro_7; 1.
 DR PRINTS; PR00734; GLHYDRASE7.
 DR ProDom; PD001821; CBD_fungal; 1.
 DR ProDom; PD186135; Glyco_hydro_7; 1.
 DR SMART; SM00236; fcBD; 1.
 DR PROSITE; PS00562; CBD_FUNGAL; 1.
 KW Carbohydrate metabolism; Cellulose degradation; Glycosidase;
 KW Hydrolase; Polysaccharide degradation; Signal; Xylan degradation.
 FT SIGNAL 1 24 Potential.
 FT CHAIN 25 529 xylanase/cellobiohydrolase.
 SQ SEQUENCE 529 AA; 55048 MW; 95232F53577B6416 CRC64;

Query Match 63.1%; Score 1730.5; DB 2; Length 529;
Best Local Similarity 63.1%; Pred. No. 2.9e-103;
Matches 323; Conservative 55; Mismatches 111; Indels 23; Gaps 9;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTATNSSTNCYDGNTWSSTL 60
| | : ||| | : | | | | : | : :: : ||| | | | : ||| | | | : | : : :
Db 26 QQIGTYTAETHPSLSWSTCKSGSCTTNSGAITLDANWRWVHGVTSTNCYTGNTWNTAI 85

Qy 61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQE 120
| : : || : | || | : || || : || || | : || | || : || | : || |
Db 86 CTDASCAQDCALDGADYSGTYGITSGNSLRLNFVTGS --NVGSRTYLMADNTHYQIF 142

Qy 121 TLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDL 180
|| | : || || || || || || : || || || || || || || : || || || || || || ||
Db 143 DLLNQEFTFTVDVSNLPCGLNGALYFVTMDADGGVSKYPNNKAGAQYGVGVYCDSQCPRDL 202

Qy 181 KFINGQANVEGWEPESSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEG 240
 ||| :|||||| | :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
 Db 203 KFIAGQANVEGWTPSTNNNSNTGIGNHGSCCAELDIWEANSISEALTPHPCDTGLTVCTA 262

Qy 297 -SGA---INRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEAAEFG-GSSFSDKGGLTQ 351
 || : | ||||||| | |::: | || : | :| || : || : || : || : ||
 Db 321 SSGSLSEIIRRYYVONGVVIPOPSSKISG1SGNVINSDECAELSAFGETASFTNHGG1KN 380

Qy 352 FKKATSGGMVLVMSLWDDYYANMLWLDSYPTNETSSTPGAVRGSCSTSSGVPAQVESQS 411
 | ||||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 381 MGSALAEAGMVLVMSLWDDYSVNMLWLDSYPTNETSSTPGAVRGSCPTTSGNPKTVESQS 439

Title: US-10-804-785-2

RESULT 15
GUX1_HUMGT
ID GUX1_HUMGT STANDARD; PRT; 525 AA.
AC P15828;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Exoglucanase I precursor (EC 3.2.1.91) (Exocellulobiohydrolase I) (1,4-beta-cellulobiohydrolase) (Beta-glucancellulobiohydrolase).
GN Name=CBH-1;
OS Humicola grisea var. thermoidea.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX NCBI_TaxID=5528;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=90175006; PubMed=2308855;
RA de Oliveira Alzevedo M., Radford A.;
RT "Sequence of cbh-1 gene of Humicola grisea var. thermoidea.";
RL Nucleic Acids Res. 18:668-668(1990).
CC -!- FUNCTION: The biological conversion of cellulose to glucose generally requires three types of hydrolytic enzymes: (1) Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2) Exocellulobiohydrolases that cut the disaccharide cellobiose from the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-glucosidases which hydrolyze the cellobiose and other short cellulose oligosaccharides to glucose.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages in cellulose and cellooligosaccharides, releasing cellobiose from the nonreducing ends of the chains.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 7 (cellulase C) family.
CC -!- SIMILARITY: Contains 1 CBM1 (fungal-type carbohydrate-binding)

CC , domain.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
DR EMBL; X17258; CAA35159.1; -; Genomic_DNA.
DR PIR; S38794; S38794.
DR HSSP; Q09431; 1GPI.
DR InterPro; IPR000254; CBD_fun.
DR InterPro; IPR001722; Glyco_hydro_7.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR PRINTS; PR00734; GLHYDRLASE7.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD186135; Glyco_hydro_7; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Carbohydrate metabolism; Cellulose degradation; Glycoprotein;
KW Glycosidase; Hydrolase; Polysaccharide degradation; Signal.
FT SIGNAL 1 18 Potential.
FT CHAIN 19 525 Exoglucanase I.
FT DOMAIN 490 525 CBM1.
FT REGION 19 467 Catalytic.
FT REGION 468 489 Linker.
FT ACT_SITE 231 231 Nucleophile (By similarity).
FT ACT_SITE 236 236 Proton donor (By similarity).
FT CARBOHYD 289 289 N-linked (GlcNAc. . .) (Potential).
FT DISULFID 497 514 By similarity.
FT DISULFID 508 524 By similarity.
SQ SEQUENCE 525 AA; 55694 MW; A6684D4CF881E090 CRC64;

Query Match 60.2%; Score 1652; DB 1; Length 525;
Best Local Similarity 57.3%; Pred. No. 3.3e-98;
Matches 294; Conservative 76; Mismatches 121; Indels 22; Gaps 7;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL 60
| ||::| :| || |::|| ||::|| | : :|:||| | : ||||| || | :::
Db 19 QQACSLTTERHPSLSWNKCTAGGQCQTVQASITLDSNWRWTHQVSGSTNCYTGNKWDTI 78

Qy 61 CPDNETCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQK-NVGARLYLMASDTTYQE 119
| | ::||:|||:||| | |||:|||:|||: | ||: | |||: ||| : |||
Db 79 CTDAKSCAQNCVVDGADYTSTYGITNGDSLKFKVTKGQHSTNVGSRTYLMGEDKYQT 138

Qy 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
| ||||||:|||:||| : |||:|||:|||:|||:|||: |||:|||:|||:|||:|||
Db 139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMADGGLSRYPGNKAGAKYGTGYCDAQCPRD 198

Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALT PHPCTTVGQEICE 239
:|||:|||:||| | :| | | | :| |||:|||:|||:||| | :|||:|||
Db 199 IKFINGEANIEGTGSTNDPNAGAGRYGTCCSEMDIWEANNMATAFTPHTCIIIGQSRCE 258

Qy 240 GDGCGBTYSNDRYYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTFETS-- 297
|| ||||||: || | |||:|||: || | | :||| | | :|||:|||:|||
Db 259 GDSCGGTYSNERVAGVCDPDGCDFNYSRQGNKTFYKGK--MTVDTTKITVVTQFLKDAN 316

Qy 298 ---GAINRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGG-SSFSDKGLTQFK 353
| | | :|||: | : : | | : | : | | : | | : | | : | |
Db 317 GDLGEIKRFYVQDGKIIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG 376

Qy 354 KATSGGMVLVMSLWDDYYANMLWLSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
| | :| | |||:|||: | :|||:|||: | : | | | : | | :|||:|||:|||
Db 377 KALAGPMVLVMSIWDDHASNMLWLSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN 435

Qy 414 AKVTFSNIKFGPIGST-----GNPSGGNPPGGNPPGTTTTRPATTGSSPGPTQS 464

Db : | ||||:|||||| | :|||| ||||:|||| :| ||
436 SNVVFSNIRFGPIGSTVAGLPAGNGGNNGNPP---PPTTTSSAPATTTASAGPKAG 492

Qy 465 HYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL 497

: |||||:||| | | || :||||

Db 493 RWQQCGGIGFTGPTQCEEPYICTKLNDWYSQCL 525

Db 138 FTLLGNEFTFDVTGLCGLNGALYFVSMDEDGGLSKYSGNKAGAKYGTGYCDSQCPRD 197
 Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
 |||||:|| ||| : | | : |||:|||:|||:|||:|||:|||:|||:|||:|||:
 Db 198 LKFINGEANNVGWTPSSNDKNAGLGNYGSCCSEMDVWEANSISAAYTPHPCTTIGQTRCE 257
 Qy 240 GDGCGGTYSDNRYGGCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLTVVTFETSGA 299
 |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 Db 258 GDDCGGTYSTDRYAGECDPDGCFNSYRMGNTFYKGK--MTVDTSKKFTVVTQFLTDSS 315
 Qy 300 -----INRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEAAEFGGSS-FSDKGLTQFK 353
 | |:|||:|||:|||:|||:|||:|||:|||:|||:
 Db 316 GNLSEIKRFYVQNGVVIPNSNSNIAGVSGNSITQAFCDAAQKTAFGDTNVFDQKGGLAQMG 375
 Qy 354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
 |||:|||:|||:|||:|||:|||:|||:|||:|||:
 Db 376 KALAQPMLVMSLWDDHAVNMLWLDSTYPTN-AAGKPGAARGTCPTTSGVPADVESQAPN 434
 Qy 414 AKVTFSNIKFGPIGST--GNPSGGNPPGNPPGTTTRPATTTGSSP-----GPTQSH 465
 :|||:|||:|||:|||:|||:|||:|||:|||:
 Db 435 SKVIYSNIRFGPIGSTVSGLPGGSNPGGSSSTTTTRPATTTSSASSGPTGGTAAH 494
 Qy 466 YGQC GGIGYSGPTVCASGTTCQVLNPYYSQCL 497
 :|||:|||:|||:|||:|||:
 Db 495 WGQC GGIGWTGPTVCASPYTCQKLNDWYYQCL 526

RESULT 134

AAB81926

ID AAB81926 standard; protein; 529 AA.

XX

AC AAB81926;

XX

DT 25-JUN-2001 (first entry)

XX

DE Acremonium cellulolyticus cellobiohydrolase 1 precursor.

XX

KW Cellobiohydrolase 1; cbh1; promoter; protein production.

XX

OS Acremonium cellulolyticus.

XX

FH Key Location/Qualifiers

FT Peptide 1. .26

FT /label= signal_peptide

FT Protein 27. .529

FT /label= mature_cellobiohydrolase

XX

PN JP2001017180-A.

XX

PD 23-JAN-2001.

XX

PF 06-JUL-1999; 99JP-00191221.

XX

PR 06-JUL-1999; 99JP-00191221.

XX

PA (MEIJ) MEIJI SEIKA KAISHA LTD.

PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX

DR WPI; 2001-294133/31.

DR N-PSDB; AAF85588.

XX

PT New promotor useful for expression of a protein.

XX

PS Disclosure; Page 12-14; 22pp; Japanese.

XX

CC The present invention provides a promoter capable of causing the expression of a gene connected downstream. It can be used for expressing

CC a protein in a large amount. The present sequence is the Acremonium
CC cellulolyticus cellobiohydrolase 1 precursor (cbh1) protein
XX
SQ Sequence 529 AA;

Query Match 63.6%; Score 1741.5; DB 4; Length 529;
Best Local Similarity 62.9%; Pred. No. 2.1e-104;
Matches 324; Conservative 52; Mismatches 110; Indels 29; Gaps 8;

Qy	1	QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL	60
Db	26	QQIGTYTAETHPSLSWSTCKSGSCTTNSGAITLDANWRWVHGVTSTNCYTGNTPNSAI	85
Qy	61	CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEF	120
Db	86	CDTDASCAQDCALDGADYSGTYGITTSGNSLRLNFTGSG--NVGSRTYLMADNTHYQIF	142
Qy	121	TLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRDL	180
Db	143	DLLNQEFTFTVDVSHLPCGLNGALYFVTMDADGGVSKYPNNKAGAQYGVGVYCDSQCPRDL	202
Qy	181	KFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEG	240
Db	203	KFIAGQANVEGWTPSSNNANTGIGNHGACCAELDIWEANSISEALTPHPCDTPGLSVCTT	262
Qy	241	DGCGGTYSDNRYGGCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTFETS---	297
Db	263	DACGGTYSSDRYAGTCDPGCDFNPyRLGVTDYGSK--TVDTTKPFTVVTFQVTNDGT	320
Qy	298	-----GAINRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGG-SSFSDKGGLTQ	351
Db	321	STGSLSEIRRYYVQNGVVIQPQSSKISGSGNVINSDYCAAEISTFGGTASFSKHGLTN	380
Qy	352	FKKATSGGMVLVMSLWDDYYANMLWLDSYPTNETSSTPGAVRGSCSTSSGVPAQVESQS	411
Db	381	MAAGMEAGMVLVMSLWDDYAVNMLWLDSYPTNAT-GTPGAARGTCATTSGDPKTVEAQS	439
Qy	412	PNAKVTFSNIKFGPIGSTGNPSGGNPPGGNPPGTTTRRPAATTGSSPGPTQS-----	464
Db	440	GSSYVTFSDIRVGPFNSTF--SGGSSTGGS---TTTASRTTTSASSTSTSSTGTGV	494
Qy	465	--HYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL	497
Db	495	AGHWGQCGGOGWTGPTTCVSGTTCTVVNPYYSQCL	529

RESULT 137
AAB47783
ID AAB47783 standard: protein: 526 AA

XX

AC AAB47783:

xx

DT 13-MAR-2002 (first entry)

BT 15 MAR 2002 (First entry)
XX

DE Chrysoporium CBH1

BB Chrysoperlam CBRI.

KW Glycosyl hydrolase; family 7; fami

RW Glycosyl hydrolase; family 7; family
KW promoter; terminator; glycosaldehyde

RW promoter, terminator; glyceraldehy-
xy

xx
OS Chrysosporium sp.

OS Chrysosporium sp.

EU Key **Location/Qualification**

FH Key Location/Qualifier
FT Partida 1-16

FT Peptide 1. .19

FT /label= signal_pep
FT Protein 22 526

FT Protein 20.526

FT /label= mature_pro
FT Missed differ

Qy 414 AKVTFNSNIKFGPIGST--GNPSG--GNPPGGNPPGTTTRP--ATTGSSPGPT---- 462
: | ||||:||||| | | | ||| ::| | :||: | |||
Db 435 SNVIFSNIRFGPIGSTVGLPDGGSGNP---NPPVSSTPVSSSTTSSGSSGPTGGTGV 491

Qy 463 QSHYGCQCGGIGYSGPTVCASGTTCQVLNPYYSQCL 497
| | ||||:| | | | ||| :|||||
Db 492 AKHYEQCGGIGFTGPTQCESPYTCTKLNDWYSQCL 526

RESULT 138

ABJ26886

ID ABJ26886 standard; protein; 529 AA.

XX

AC ABJ26886;

XX

DT 08-MAY-2003 (first entry)

XX

DE Cellobiohydrolase I activity protein SEQ ID No 4.

XX

KW Cellobiohydrolase; enzyme; DNA shuffling; ethanol; biomass;
KW cellobiohydrolase I; EC 3.2.1.91.

XX

OS Chaetomium thermophilum.

XX

PN WO2003000941-A2.

XX

PD 03-JAN-2003.

XX

PF 26-JUN-2002; 2002WO-DK000429.

XX

PR 26-JUN-2001; 2001DK-00001000.

XX

PA (NOVO) NOVOZYMES AS.

XX

PI Lange L, Wu W, Aubert D, Landvik S, Schnorr KM, Clausen IG;

XX

DR WPI; 2003-278244/27.

DR N-PSDB; ABT23504.

XX

PT New polypeptide with cellobiohydrolase I activity, useful in producing
PT ethanol from biomass.

XX

PS Claim 4; Page 115-117; 199pp; English.

XX

CC The invention relates to a novel polypeptide comprising: part of any of
CC 21 amino acid sequences; an amino acid sequence at least 70% identical to
CC a polypeptide encoded by a cellobiohydrolase gene; an amino acid sequence
CC at least 80% identical to the polypeptide encoded by 21 nucleotide
CC sequences; a polypeptide encoded by a nucleotide sequence which
CC hybridises with a probe selected from complementary strands of 55
CC nucleotide sequences; or a fragment of the aforementioned structures. The
CC polynucleotides of the invention are useful in a method of DNA shuffling.
CC The polypeptides are useful in a method for producing ethanol from
CC biomass comprising contacting the biomass with the polypeptides. This
CC sequence represents a protein with cellobiohydrolase I activity of the
CC invention

XX

SQ Sequence 529 AA;

Query Match 61.6%; Score 1689; DB 6; Length 529;
Best Local Similarity 59.3%; Pred. No. 5.3e-101;
Matches 305; Conservative 76; Mismatches 113; Indels 20; Gaps 7;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL 60
| ||:| :||| | ||::|:| | :| :||| | ||| :| ||| | | | :::
Db 19 QQACSLTTETHPRLTWKRCTSGGNCTVNGAVTIDANWRWHTHTVSGSTNCYTGNEWDTI 78

Qy	61	CPDNEXCAKNCLDGAAAYASTYGVTTSGNSLSIGFVTQSAQ-KNVGARLYLMASDTTYQE	119
: : : : : : : : : : : : : :			
Db	79	CSDGKSCAQTCVDGADYSSTYGITSGDSLNLKFVTKHQYGTNVGSRVYLMENDTKYQM	138
Qy	120	FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRD	179
:			
Db	139	FELLGNEFTFDVDVSNLGCGLNGALYFVSMADGGMSKYSGNKAGAKYGTGYCDAQCPRD	198
Qy	180	LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE	239
:			
Db	199	LKFINGEANIEENWTPSTNDANAGFGRYGSCCSEMDIWEANNMATAFTPHTCIIQGSRCE	258
Qy	240	GDGCGGTYSNDNRYGGTCDPDGCDWNPYRLGNTSFYPGPSSFTLDTTKKLTVVTQFETSGA	299
:			
Db	259	GNSCGGTYSERYAGVCDPDGDFNAYRQGDKTFYGKG--MTVDTTKKMTVVTQFHKNSA	316
Qy	300	-----INRYYVQNQVTFQQPNAELGSYSGNELNDDYCTAEAEFGG-SSFSDKGGLTQFK	353
: : :: : : ::			
Db	317	GVLSEIKRFYVQDGKVIANAESKIPGNPGNSITQEWCDAQKVAFGDIDDFNRKGGMAQMS	376
Qy	354	KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN	413
: :			
Db	377	KALEGPMLVMSVWDDHYANMLWLDSTYPIDK-AGTPGAERGACPTTSGVPAEIEAQVPN	435
Qy	414	AKVTFSNIKFGPIGSTGNPSGGNPPGG----NPPGTTTRRPATTTGSSP----GPTQ	463
: :			
Db	436	SNVIFSNIRFGPIGSTVPGLDGSTPSNPTATVAPPSTTSVRSSSTQISTPTSQPGGCTT	495
Qy	464	SHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL	497
:			
Db	496	QKWGQCGGIGYTGCTNCVAGTTCTELNPWYSQCL	529

RESULT 140
ABB05058
ID ABB05058 standard; protein; 526 AA.
XX
AC ABB05058;
XX
DT 11-SEP-2003 (revised)
DT 27-MAR-2002 (first entry)
XX
DE Trichoderma reesei cellobiohydrolase I (CBH1) 55kD (family 7) protein.
XX
KW Trichoderma reesei; filamentous fungi; phenotype; characterisation;
KW fermentation; screening; morphology; cellobiohydrolase I; CBH1.
XX
OS Hypocrea jecorina.
XX
FH Key Location/Qualifiers
FT Peptide 1. .19
FT /label= signal
FT Protein 20. .526
FT /label= cellobiohydrolase_I
FT Misc-difference 249
FT /label= unknown
FT /note= "encoded by NCC"
FT Misc-difference 365
FT /label= unknown
FT /note= "encoded by TTN"
XX
PN WO200125468-A1.
XX
PD 12-APR-2001.
XX
PF 13-APR-2000; 2000WO-US010199.

xx
PP 06 OCT 1988 200-618

PA (EMAL/) EMALFARB M A.

PT Emalfarb MA:

DB WBT: 2001-281733/38

DB N-PSDB: ABA92723

PT Expressing heterologous proteins encoded by a library of DNA vectors, PT involves stably transforming mutant filamentous fungus with the vectors PT and culturing transformed fungi for expressing heterologous proteins.

PS Disclosure; Page 66-69; 85pp; English.

The present invention describes a method of expressing a number of proteins encoded by a library of DNA vectors (I). The method involves stably transforming a mutant filamentous fungus (II) with (I) so as to introduce into each of a number of individual fungi, at least one heterologous protein-encoding nucleic acid sequence (III), and culturing the transformed mutant filamentous fungi for the expression of heterologous proteins encoded by (III). (I) comprises a number of different vectors, each comprising a different protein-encoding nucleic acid sequence being operably linked to an expression regulating region and optionally a secretion signal encoding sequence. (II) has a phenotype characterised by a culture viscosity, when cultured in suspension, of less than 200 cP at the end of fermentation when grown with adequate nutrients under optimal or near-optimal conditions. The method is useful for expressing large quantities of heterologous proteins that are useful for isolation, characterisation and application testing, and also for commercial production of proteins. The mutant filamentous fungi obtained by the method are suitable for high-throughput screening techniques owing to their unique morphology and very low viscosity of their cultures. The present sequence represents the *Trichoderma reesei* cellobiohydrolase I (CBH1) 55kD (family 7) protein, which is given in the exemplification of the present invention. (Updated on 11-SEP-2003 to standardise OS field)

SQ Sequence 526 AA;

Query Match 61.4%; Score 1681; DB 4; Length 526;
Best Local Similarity 60.4%; Pred. No. 1.7e-100;
Matches 311; Conservative 68; Mismatches 112; Indels 24; Gaps 10;

Qy 61 CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA-QKNVGARLYLMASDFTTYQE 119
| | :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 78 CSDGPSCASKCCIDGADYSSTYGITSGNSLNLKFKVTKGOYSTNIGSRTYLMESDTKYOM 137

Qy 120 FTLLGNEFSFDVDSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
| :||||||:|||||| | :||||||:|||||| | :|||||| | :|||||| | :|||||| |
Pb 128 FOLLGNEETEDVDSNLCCGLNGALYEVGMDAPCGMSKYGCHKAGAKYCTGYCDSQCPRD 177

Qy	180 LKFINGQANVEGWEPSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239 : : : : : : ::: :
Db	198 LKFINGEANVENWQSSTNDANAGTGYGSCCSEMDVWEANNMAAFTPHPCXVIGQRCE 257

Qy 240 GDGCGGTYSNDRYGGTCDPDGCDWNPYRLGNTSFYGPGSSFTLDTKKLTVVTQFETSGA 299
|| |||||| :|| | ||||||:| || || :||| | ||:||||:||||| : |
Db 258 GDSCGGTYSTDYAGICDPDGCFNSYRQGNKTFYGKG--MTVDTKKITVVTQFLKNSA 315

QY 300 -----INRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGGSSF-SDKGGLTQFK 353
 | : | : | : | : : | : | : | : | : | : | : | : | : | : |
 Db 316 GELSEIKRFYVQNGKVIPNSESTIPGVEGNSITQDWCDRQKAAGFDVTDXQDKGGMVQMG 375

Qy 354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
|| :| |||||:|||: |||||:| : | ||| ||:| |:||||:||::|||
Db 376 KALAGPMVLVMSIWDDHAVNMLWLDSTWPI-DGAGKPGAERGACPTTSGVPAEVEAEAPN 434

Qy 414 AKVTFSNIKFGPIGST--GNPSG--GNPPGGNPPGTTTRP--ATTGSSPGPT---- 462
: | ||||:||| | | | | | | | ::| | :||: | |||
Db 435 SNVIFSNIRFGPIGSTVSGLPDGGSGNP---NPPVSSTPVPSSTSSGSSGPTGGTGV 491

Qy 463 QSHYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL 497
|| | ||||:||| | | | | | :|||
Db 492 AKHYEQCGGIGFTGPTQCESPYTCTKLNDWYSQCL 526

RESULT 141

ABW00703

ID ABW00703 standard; protein; 526 AA.

XX

AC ABW00703;

XX

DT 15-JAN-2004 (first entry)

XX

DE Chrysosporium lucknowense cellobiohydrolase (CBH1) protein.

XX

KW Mutant Chrysosporium strain; fungal enzyme; metabolite; organic acid;
KW antibiotic; cellobiohydrolase; CBH1.

XX

OS Chrysosporium lucknowense.

XX

FH Key Location/Qualifiers

FT Peptide 1. .20

FT /label= Signal_peptide

FT Protein 21. .526

FT /note= "Mature CBH1 protein"

FT Misc-difference 137

FT /note= "Encoded by AGTAAGTCCTCTCGCACCCGGCCGCCGGAGATGAT
GGCGCCCAGCCCGCTGACGCGAATGACACAGTG"

FT Misc-difference 249

FT /note= "Encoded by ACC"

FT Misc-difference 365

FT /note= "Encoded by TTN"

FT Domain 495. .526

FT /note= "CBD domain"

XX

PN US6573086-B1.

XX

PD 03-JUN-2003.

XX

PF 13-APR-2000; 2000US-00548938.

XX

PR 06-OCT-1998; 98WO-EP006496.

PR 06-OCT-1999; 99WO-NL000618.

XX

PA (DYAD-) DYADIC INT INC.

XX

PI Emalfrab MA, Burlingame RP, Olson PT, Sinitsyn AP, Parriche M;
PI Bousson JC, Pynnonen CM, Punt PJ, Van Zeijl CMJ;

XX

DR WPI; 2003-764575/72.

DR N-PSDB; AAD61474.

XX

PT New mutant Chrysosporium strain expressing a heterologous polypeptide, or
PT overexpressing a homologous polypeptide, at a high level, useful for
PT production of e.g. enzymes, primary metabolites, and antibiotics.

XX

PS Disclosure; Col 43-44; Opp; English.

XX

The invention relates to a mutant *Chrysosporium* strain comprising a nucleic acid encoding a polypeptide of interest, linked to an expression-regulating region chosen from promoter sequences associated with cellulase, xylanase or glyceraldehyde-3-phosphate dehydrogenase (*gpdA*) expression and optionally to a secretion signal sequence, where the mutant strain expresses the polypeptide at a higher level than a non-mutant strain under same conditions. The invention is useful for producing polypeptides such as carbohydrate-degrading enzymes, proteases, lipases, esterases, other hydrolases, oxidoreductases and transferases. The invention is also useful for producing fungal enzymes allowing production or overproduction of primary metabolites, organic acids, secondary metabolites or antibiotics. The present sequence is *Chrysosporium lucknowense* cellobiohydrolase (CBH1) protein

XX
SQ Sequence 526 AA;

Query Match 61.4%; Score 1681; DB 7; Length 526;
Best Local Similarity 60.4%; Pred. No. 1.7e-100;
Matches 311; Conservative 68; Mismatches 112; Indels 24; Gaps 10;

RESULT 142

ABJ26904

ID ABJ26904 standard; protein; 525 AA.

xx

AC ABJ26904;

xx

DT 08-MAY-2003 (first entry)

xx

DE Cellobiohydrolase I activity protein SEQ ID No 60.

KW Cellobiohydrolase; enzyme; DNA shuffling; ethanol; biomass;
KW cellobiohydrolase I; EC 3.2.1.91.
XX
OS *Scytalidium thermophilum*.
XX
PN WO2003000941-A2.
XX
PD 03-JAN-2003.
XX
PF 26-JUN-2002; 2002WO-DK000429.
XX
PR 26-JUN-2001; 2001DK-00001000.
XX
PA (NOVO) NOVOZYMES AS.
XX
PI Lange L, Wu W, Aubert D, Landvik S, Schnorr KM, Clausen IG;
XX
DR WPI; 2003-278244/27.
DR N-PSDB; ABT23542.
XX
PT New polypeptide with cellobiohydrolase I activity, useful in producing
PT ethanol from biomass.
XX
PS Claim 4; Page 191-192; 199pp; English.
XX
CC The invention relates to a novel polypeptide comprising: part of any of
CC 21 amino acid sequences; an amino acid sequence at least 70% identical to
CC a polypeptide encoded by a cellobiohydrolase gene; an amino acid sequence
CC at least 80% identical to the polypeptide encoded by 21 nucleotide
CC sequences; a polypeptide encoded by a nucleotide sequence which
CC hybridises with a probe selected from complementary strands of 55
CC nucleotide sequences; or a fragment of the aforementioned structures. The
CC polynucleotides of the invention are useful in a method of DNA shuffling.
CC The polypeptides are useful in a method for producing ethanol from
CC biomass comprising contacting the biomass with the polypeptides. This
CC sequence represents a protein with cellobiohydrolase I activity of the
CC invention
XX
SQ Sequence 525 AA;

Query Match 60.8%; Score 1666; DB 6; Length 525;
Best Local Similarity 57.5%; Pred. No. 1.6e-99;
Matches 295; Conservative 78; Mismatches 118; Indels 22; Gaps 7;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL 60
| |::| :| || |::|::|| | | |::|:||||| : ||||| || | :::
Db 19 QQACSLTTERHPSLSWKCTAGGQCQTQASITLDSNWRWTHQVSGSTNCYTGNKWDTSI 78

Qy 61 CPDNEXCAKNCCLDGAAYASTYGVTSGNSLSIGFVTQSAQK-NVGARLYLMASDTTYQE 119
| | ::|:|:|||:||| | |||:|||:|||:|:|||: | |||: | ||| : |||
Db 79 CTDAKSCAQNCVVDGADYTSTYGITNGDSLSSLKFVTKGQHSTNVGSRTYLMGEDKYQT 138

Qy 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
| ||||||:||||| : |||||||||||||||:|:||| | |||||||||||:|||||
Db 139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMDADGGLSRYPGNKAGAKYGTGYCDAQCPRD 198

Qy 180 LKFINGQANVEGWEPPSNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
:|||||:|||:||| |::| | | | :|:|||||:|||: | ||||| :||| |||
Db 199 IKFINGEANIEGTGSTNDPNAGAGRGTCCSEMDIWEANNMATAFTPHPCTIIGQSRCE 258

Qy 240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTFETS-- 297
| | |||||: | | | |||||:| | | | :| | | | :|:|||||:|||||
Db 259 GDSCGGTYSNERYAGVCDPDGCDFNYSRQGNKTFYKGK--MTVDTTKKITVVTQFLKDAN 316

Qy 298 ---GAINRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGG-SSFSDKGGLTQFK 353
| : |:|||:| : : | | : |: | : | | : | |||: |
Db 317 GDLGEVKRFYVQDGKIIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG 376

Qy 240 GDGCGGTYSNDRYGGCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTFETSGA 299
|| || || | . || | || || || : | || | . || | | : || || || : |
Db 258 GDSCGGTYSTDYAGICDPDGDFNSYRQGNKTFYGKG--MTVDTKKITVVTFQLKNSA 315

Qy 300 -----INRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEAEFGGSSF-SDKGLTQFK 353
| | : || || | . : : || : | : || : | : | || : |
Db 316 GELSEIKRFYVQNGKVIPNSESTIPGVEGNSITQDWCDRQKAAGDVTDXQDKGGMVQMG 375

Qy 354 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
| | : | || || || : | | : | || || || : | : || : | : ||
Db 376 KALAGPMVLVMSIWDDHAVNMLWLDSTWPI-DGAGKPGAERGACPTTSGVPAEVEAEAPN 434

Qy 414 AKVTFSNIKFGPIGST--GNPSG--GNPPGGNPPGTTTRRP--ATTGSSPGPT---- 462
: | || || : | || || | | | | | || :: : | | : | : | ||
Db 435 SNVIFSNIRFGPIGSTVSLPDGGSGNP--NPPVSSSTPVSSSTSSGSSGPTGGTGV 491

Qy 463 QSHYGCQCGGIGYSGPTVCASGTTCQVLNPYYSQCL 497
| | || || | : | : | | | | || : | : ||
Db 492 AKHYEQCGGIGFTGPTQCESPYTCTKLNDWYSQCL 526

RESULT 10

US-08-676-166A-3

; Sequence 3, Application US/08676166A

; Patent No. 5955270

; GENERAL INFORMATION:

; APPLICANT: Radford, Alan

; APPLICANT: Parish, John H.

; TITLE OF INVENTION: EXPLOITATION OF THE CELLULASE COMPLEX OF

; TITLE OF INVENTION: NEUROSPORA

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: David A. Jackson, Esq.

; STREET: 411 Hackensack Ave, Continental Plaza, 4th

; STREET: Floor

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/676,166A

; FILING DATE: 15-JUL-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 1321-1-002

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-487-5800

; TELEFAX: 201-343-1684

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 525 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; HYPOTHETICAL: NO

; ORIGINAL SOURCE:

; ORGANISM: H. grisea

US-08-676-166A-3

Query Match 60.5%; Score 1658; DB 1; Length 525;
Best Local Similarity 57.5%; Pred. No. 1.4e-119;
Matches 295; Conservative 77; Mismatches 119; Indels 22; Gaps 7;

RESULT 16

US-08-676-166A-2

; Sequence 2, Application US/08676166A

Patent No. 5955270

GENERAL INFORMATION:

APPLICANT: Radford, Alan

APPLICANT: Parish, John H.

TITLE OF INVENTION: EXPLOITATION

TITLE OF INVENTION: NEUROSPORA

NUMBER OF SEQUENCES: 7

NUMBER OF SEQUENCES: ;
CORRESPONDENCE ADDRESS: ;

ADDRESSEE: David A. Jackson Esq.

ADDRESSEE: David
STREET: 411 Hac

STREET: Floor

STREET: FLOOR
CITY: Hackensack

CITY: Hackensack
STATE: New Jersey

STATE: NEW J
COUNTRY: USA

COUNTRY: U
ZIP: 07601

ZIP: 07601
COMPUTER READABLE FORM

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS/MS DOS

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: BASIC

SOFTWARE: PatentIn Re.
CURRENT PRACTICE IN PATENT

CURRENT APPLICATION DATA:

FILING DATE: 15-JUL-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 1321-1-002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
FORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 516 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-676-166A-2

Query Match 57.0%; Score 1561; DB 1; Length 516;
Best Local Similarity 57.5%; Pred. No. 4.2e-112;
Matches 294; Conservative 62; Mismatches 129; Indels 26; Gaps 10;

RESULT 19

US-09-329-350-35

; Sequence 35, Application US/09329350

; Patent No. 6184019

; GENERAL INFORMATION:
; APPLICANT: Miettinen-Oinonen, Arja
; APPLICANT: Londesborough, John
; APPLICANT: Vehmaanper , Jari
; APPLICANT: Haakana, Heli

APPLICANT: M ntyl , Arja
APPLICANT: Lantto, Raija
APPLICANT: Elovainio, Minna
APPLICANT: Joutsjoki, Vesa
APPLICANT: Paloheimo, Marja
APPLICANT: Suominen, Pirkko
TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/329,350
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/841,636
FILING DATE: 30-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/005,335
FILING DATE: 17-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,926
FILING DATE: 04-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/020,840
FILING DATE: 28-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/732,181
FILING DATE: 16-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FI96/00550
FILING DATE: 17-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Shea Jr., Timothy
REGISTRATION NUMBER: 41,306
REFERENCE/DOCKET NUMBER: 1716.0510006/MAC/TJS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)371-2600
TELEFAX: (202)371-2540
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 452 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Melanocarpus albomyces
STRAIN: ALKO4237
FEATURE:
NAME/KEY: Protein
LOCATION: 1..452
OTHER INFORMATION: /label= 50K-cellulase-B

US-09-329-350-35

RESULT 20

11S-08-841-636A-35

: Sequence 35. Application US/08841636A

Patent No. 6723549

GENERAL INFORMATION:

; APPLICANT: Miettinen-Oinonen, Arja
; APPLICANT: Londenborough, John
; APPLICANT: Vehmaanper , Jari
; APPLICANT: Haakana, Heli
; APPLICANT: M ntyl , Arja
; APPLICANT: Lantto, Raija
; APPLICANT: Elovainio, Minna
; APPLICANT: Joutsjoki, Vesa
; APPLICANT: Paloheimo, Marja
; APPLICANT: Suominen Pirkko

APPLICANT: SUOMINEN, PIKKI
TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND
TITLE OF INVENTION: USES THEREOF

1 TITLE OF INVENTION: USES THEREOF
1 NUMBER OF SEQUENCES: 45

NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:

ADDRESSEE: Stores Kessler Goldstein & Fox P. I. L. S.

STREET: 1100 New York Avenue, N.W., Suite 600

STREET: 1100 New
CITY: Washington

CITY: washin
STATE: P.C.

STATE: D.C.
COUNTRY: USA

COUNTRY: US
ZIP: 2000E

; ZIP: 20003
COMPUTER READABLE FORM

COMPUTER READABLE FORM:
MEDIUM TYPE - Punched - 3 1/2 inch

; MEDIUM TYPE: Diskette, 3.50 COMPUTER: IBM PC

; COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS/VG DOS

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Radio-TEC Model 1000 Version 1.0

**SOFTWARE: PatentIn Rel
CURRENT APPLICATION DATE**

Db 326 DGRKIEIPPPTE-GMPNSSEITPELCSTMFDFNDRNRFEVGGFEQLNNALRPMVLV 384
 Qy 364 MSLWDDYYANMLWLDSITYPTNETSSTPGAVRGSCSTSSGVPQAQVESQSPNAKVTFSNIKF 423
 ||:|||:||||| || | || || | || |||:||:| :||:| :|||:
 Db 385 MSIWDDHYANMLWLDSIYPP-EKEGQPGAARGDCPTDSGVPAEVEAQFPDAQVVWSNIRF 443
 Qy 424 GPIGST 429
 |||||
 Db 444 GPIGST 449

RESULT 32

US-08-709-979A-1

; Sequence 1, Application US/08709979A

; Patent No. 5912157

; GENERAL INFORMATION:

; APPLICANT: Claus von der Osten
 ; APPLICANT: Martin Sch lein
 ; TITLE OF INVENTION: No. 5912157el Alkaline Cellulases
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 5912157o No. 5912157disk of No. 5912157th America, Inc.
 ; STREET: 405 Lexington Avenue, 64th Floor
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States of America
 ; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/709,979A
 ; FILING DATE: 09-SEP-1996
 ; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.
 ; REGISTRATION NUMBER: 33,728
 ; REFERENCE/DOCKET NUMBER: 4160.404-US

TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123
 ; TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

; LENGTH: 456 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein

US-08-709-979A-1

Query Match 28.0%; Score 767.5; DB 1; Length 456;
 Best Local Similarity 36.5%; Pred. No. 5.3e-51;
 Matches 173; Conservative 71; Mismatches 161; Indels 69; Gaps 16;

Qy 9 ETPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNEXC 67
 | || :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
 Db 28 EVHPQITTYRCTKADGCEEKTNYIVLDALSHPVHQVDNPYNCGDWQKPNETACPDLESC 87

Qy 68 AKNCCLDGAAAYASTYGVTTSGNSLSIGFVTQSAQKN-VGARLYLMASDTT---YQEFTLL 123
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
 Db 88 ARNCIMDPVSDYGRHGSTDGTSLRL---KQLVGGNVSPRVYLL--DETKERYEMLKLT 142

Qy 124 GNEFSFDVDVSQPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PRDL 180
 ||||:|||| :||||:| ||| ||| | :| | ||| :||||:||| |||
 Db 143 GNEFTFDVDAKLPCGMNSALYLSMADATGARSE--LNPGGATFGTGYCDAQCYVTP--- 197

Qy 181 KFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEG 240
 |||| | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 Db 198 -FINGLGNIE-----GKGACCNEMDIWEANARAQHIAFPCKAGPYLCEG 242
 |:
 Qy 241 DGCGGTYSNDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTQFETSG-- 298
 | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
 Db 243 AEC-----EFDGVCDKNGCAWNPYRVNVTDYYGEGAEFRVDTTRPFSVVTQFRAGGDA 295
 |:
 Qy 299 -----AINRYYQNGVTFQQPNAEL-GSYSGNELNDDYCTAEEAEFGGSSFSDKGLTQ 351
 : | | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 Db 296 GGGKLESIYRLFVQDGRVIESYVVDKPGLPPTDRMTDEFCAAT---GAARFTELGAMEA 351
 |:
 Qy 352 FKKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQS 411
 | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
 Db 352 MGDALTRGMVLALSIWWSEGDNMNWLDS-----GEAGPCDPDEGNPSNIIRVQ 399
 |:
 Qy 412 PNAKVTFSNIKFGPIGSTGNPSGNPPGGNPPGTTTRRPATTGSSPGPTQSH 465
 | : | | | : | : | : | : | : | : | : | : | : | : | : | : | : |
 Db 400 PDPEVVFSNLRWGEIGST-YESAVDGPVGKGKGKGKAPA---GDGNGKEKSN 449

RESULT 33

US-08-709-974A-11

; Sequence 11, Application US/08709974A

; Patent No. 6117664

; GENERAL INFORMATION:

; APPLICANT: Sch lein, Martin

; APPLICANT: Rosholm, Peter

; APPLICANT: Nielsen, Jack Bech

; APPLICANT: Hansen, Svend Aage

; APPLICANT: von der Osten, Claus

; TITLE OF INVENTION: No. 6117664el Alkaline Cellulases

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6117664o No. 6117664disk of No. 6117664th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/709,974A

; FILING DATE: 09-SEP-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Gregg, Valeta

; REGISTRATION NUMBER: 35,127y

; REFERENCE/DOCKET NUMBER: 4160.414-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 456 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-709-974A-11

Query Match

27.8%; Score 762.5; DB 2; Length 456;

Best Local Similarity 36.3%; Pred. No. 1.3e-50;
Matches 172; Conservative 72; Mismatches 161; Indels 69; Gaps 16;

Qy 9 ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNEXC 67
| | | : | : | : | : | : | : | : | : | : | : | : |
Db 28 EVHPQITTYRCTKADGCEEKTNYIVLDALSHPVHQVDNPYNCGDWGQKPNETACPDLESC 87

Qy 68 AKNCLLDGAAYASTYGVTSGNSLSIGFVTQSAQKN-VGARLYLMASDTT---YQEFTLL 123
| : | | : | : | : | : | | : | | | | : | | | : |
Db 88 ARNCIMDPVSDYGRHGSTDGTSRL---KQLVGGNVSPRVYLL--DETKERYEMLKLT 142

Qy 124 GNEFSFDVDVSQQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC---PRDL 180
| | | : | | | : | | | | | | | | | | | | | | | |
Db 143 GNEFTFDVDAKLPCGMNSALYLSEMDATGARSE--LNPGGATFGTGYCDAQCYVTP--- 197

Qy 181 KFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEG 240
| | | | : | | | : | : | | | | | | | | | | | | |
Db 198 -FINGLGNIE-----GKGACCNEMDIWEANARAQHIAAPHPCSKAGPYLCEG 242

Qy 241 DGCGGTYSNDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTFETSG-- 298
| | | : | | | : | | | | | | | | | | | | | | | |
Db 243 AEC-----EFDGVCDKNGCAWNPYRVNVTDYEGAEFRVDTTRPPSVVTQFRAGGDA 295

Qy 299 -----AINRYYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEEAEFGGSSFSDKGLTQ 351
| | | : | | : | : | | | : | | | | | | | | | | : |
Db 296 GGGKLESIYRLFVQDGRVIESYVVDKPGPPTDRMTDEFCAAT---GAARFTELGAMEA 351

Qy 352 FKKATSGGMVLVMSLWDDYYANMLWLDSITYPTNETSSTPGAVRGSCSTSSGVPAQVESQS 411
| : | | | : | | : | | | | | | | | | | | | | | : |
Db 352 MGDALTRGMVLALSIWWSEGNDMNWLDS-----GEAGPCDPDEGNPSNIIRVQ 399

Qy 412 PNAKVTFSNIKFGPIGSTGNPSGGNPPGNPPGTTTRPATTTGSSPGPTQSH 465
| : | | | : | | | | | | | | | | | | | | | | | | : |
Db 400 PDPEVVFSNLRWGEIGST-YESAVDGPVGKGKGKGKAPA---GDGNGKEKSN 449

RESULT 34

US-09-329-350-33

; Sequence 33, Application US/09329350

; Patent No. 6184019

; GENERAL INFORMATION:

; APPLICANT: Miettinen-Oinonen, Arja
; APPLICANT: Londesborough, John
; APPLICANT: Vehmaanper , Jari
; APPLICANT: Haakana, Heli
; APPLICANT: M ntyl , Arja
; APPLICANT: Lantto, Raija
; APPLICANT: Elovainio, Minna
; APPLICANT: Joutsjoki, Vesa
; APPLICANT: Paloheimo, Marja
; APPLICANT: Suominen, Pirkko

; TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND

; TITLE OF INVENTION: USES THEREOF

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

Qy 299 AINRYYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEEAEFGGSSFSDKGLTQFKKATS 357
 :|:| |||:| : . | : |||:| | | : | | | | :
Db 299 SIHRLYVQDGKVIESYVVDAPGLPRTDSLNEFCAAT---GAARYLDLGGTAGMDAMT 354
 /
Qy 358 GGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVT 417
 |||| |||:| | | ||| | | | : | : |||
Db 355 RGMVLAMSIWWDESGFMNWLD-----GEAGPCLPDEGDPKNIVKVEPSPEVT 402
Qy 418 FSNIKFGPIGST 429
 :||:::| ||||
Db 403 YSNLRWGEIGST 414

RESULT 35

US-08-841-636A-33

; Sequence 33, Application US/08841636A

; Patent No. 6723549

; GENERAL INFORMATION:

; APPLICANT: Miettinen-Oinonen, Arja
; APPLICANT: Londesborough, John
; APPLICANT: Vehmaanper , Jari
; APPLICANT: Haakana, Heli
; APPLICANT: M ntyl , Arja
; APPLICANT: Lantto, Raija
; APPLICANT: Elovainio, Minna
; APPLICANT: Joutsjoki, Vesa
; APPLICANT: Paloheimo, Marja
; APPLICANT: Suominen, Pirkko

; TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND
; TITLE OF INVENTION: USES THEREOF

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/841,636A
; FILING DATE: 30-APR-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/005,335
; FILING DATE: 17-OCT-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/007,926
; FILING DATE: 04-DEC-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/020,840
; FILING DATE: 28-JUN-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/732,181
; FILING DATE: 16-OCT-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/FI96/00550
; FILING DATE: 17-OCT-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Timothy J. Shea, Jr.

; REGISTRATION NUMBER: 41,306

; REFERENCE/DOCKET NUMBER: 1716.0510005/MAC/TJS

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6117664o No. 6117664disk of No. 6117664th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6401

COMPUTER READABLE FORM

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/709,974A
FILING DATE: 09-SEP-1996
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Gregg, Valeta
REGISTRATION NUMBER: 35,127Y
REFERENCE/DOCKET NUMBER: 4160.414-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:
LENGTH: 409 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-709-974A-3

Query Match 27.4%; Score 750.5; DB 2; Length 409;
Best Local Similarity 38.3%; Pred. No. 9.3e-50;
Matches 171; Conservative 51; Mismatches 160; Indels 65; Gaps 15;

Qy 9 ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNE 67
| | | | : | : | : | | : | | | | : | | | : | | | : | | | : | | | : |
Db 8 EOHPKLETYRCTKASGCKKOTNYIVADAG---THGIRRSAGCGDWGOKPNATACPDEASC 64

Qy 122 LLGNEFSFDVDVSQPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC--PR 178
| | |||||:: : |||:||||| | ||| | : ||| ||| |||:||| |

Qy 179 DLKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEIC 238

Qy 239 EGDGCGGTYSNDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTQFETSG 298

Qy 299 -----AINRYVQNGVTQQPNAEL-GSYSGNELNDDYCTAEEAEFGGSSFSDKGGLTQF 352

Qy 353 KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412

Db 331 GDAMSRGMVLAMSVWSEGDFMAWLQ----
Qy 413 NAKVTFNSNIKFGPIGSTGNPSGGNPPG 439

RESULT 37
US-09-069-632-2
; Sequence 2, Application US/09069632
; Patent No. 6261828
; GENERAL INFORMATION:
; APPLICANT: Lund, Henrik
; TITLE OF INVENTION: A Process For Combined Desizing
; TITLE OF INVENTION: And Stone-Washing of Dyed Denim
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6261828o No. 6261828disk of No. 6261828th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/069,632
; FILING DATE: 29-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK96/00469
; FILING DATE: 15-NOV-1996
; APPLICATION NUMBER: 1278/95
; FILING DATE: 15-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gregg, Valeta
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 4588.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 409 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-09-069-632-2

Query Match 27.4%; Score 750.5; DB 2; Length 409;
Best Local Similarity 38.3%; Pred. No. 9.3e-50;
Matches 171; Conservative 51; Mismatches 160; Indels 65; Gaps 15;

Qy 9 ETHPPLTQWKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNEXC 67
| | | | : | : | : | || | | | | | | | | : | || | : | | | : | | | | : | | | | : |
Db 8 EQHPKLETYRCTKASGCKKQTNYIVADAG--IHGIRRSAGCGDWGQKPNATACPDEASC 64

Qy 68 AKNCLLDGA---AYASTYGVTTSGNSLSIGFVTQSAQKN--VGARLYLMASD-TTYQEFT 121
 ||||| | | | : | : ||||| | : | | | | | | : || : : | : | : | : | : |
 Db 65 AKNCILSGMDSNAYKNA-GITTSGNKRLR---OOLINNOVLSPRVYLLNEENKKYEMLH 119

Qy 122 LLGNEFSFDVDSLQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC--PR 178
| | | |||:: : |||: |||| | | ||| : ||| ||| |||: || |
Db 120 LTGTEFSFDVEMEKLPCGMNGALYLSEMPDGKTSRNSKAGAYYGAGYCDAOCYVTP- 178

Qy 179 DLKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVQGEIC 238
||||| |::: | | | |:::|:| | | | | | : : | | | |: | | |

; MOLECULE TYPE: protein
US-08-361-920-25

RESULT 39

US-08-479-939-25

; Sequence 25, Application US/08479939

; Patent No. 5686593

; GENERAL INFORMATION:

; APPLICANT: Woeldike, Helle F.

; APPLICANT: Hagen, Frederick

; APPLICANT: Hjort, Carsten M.

; APPLICANT: Sven, Hastrup

; TITLE OF INVENTION: An Enzyme

; TITLE OF INVENTION: or Hemic

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 56865930 NO

; STREET: 405 Lexington Avenue
CITY: New York

; CITY: New York
STATE: New York

STATE: New York
COUNTRY: United States of

COUNTRY: United States of
ZIP: 10134-6201

ZIP: 10174-8201

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/M

SOFTWARE: PatentIn Release

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/361,920
FILING DATE: 22-DEC-1994
APPLICATION NUMBER: US 07/940,860
FILING DATE: 28-OCT-1992
APPLICATION NUMBER: DK 1158/90
FILING DATE: 09-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00124
FILING DATE: 08-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3435.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-0298
FORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 427 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-479-939-25

Query Match 27.4%; Score 750.5; DB 1; Length 427;
Best Local Similarity 38.3%; Pred. No. 9.9e-50;
Matches 171; Conservative 51; Mismatches 160; Indels 65; Gaps 15;

Qy 9 ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNEXC 67
| | | | : | : | : || : | | | | | | | | | : | | | | : |
Pb 26 EOHPKLETYRCKASGCKKOTNYIVADAG---IHGIPRSAGCGDWGOKPRNATACPDEASC 82

Qy	68 AKNCCLDGA---AYASTYGVTTSGNLSIGFVTQSAQKN--VGARLYLMASD-TTYQEFT 121
	: : : : : : :
Db	83 AKNCILSGMDSNAYKNA-GITTSGNKRLR---OOLINNOVSPRVYLLNEENKKKYEMLO 137

Qy 122 LLGNEFSFDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC--PR 178
 | | | | | | : : | | | | | | | | | | | | | | | | | | | | | | | |
 Db 138 LTGTEFSFDVEMEKLPCGMNGALYLSEMPQDGKSTSRSNKA G A Y Y G A G Y C D A Q C Y V T P - 196

Qy	179 DLKFINGQANVEGWEPPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEIC 238
	:: : : : : :
Db	197 ---FINGVGNIK-----GQGVCCNELDIWEANSRATHIAPHPCSKPGLYGC 239

Qy 299 -----AINRYYVQNGVTQQPNAEL-GSYSGNELNDDYCTAEEAEFGSSFSDKGGLTQF 352
 ::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
 Db 293 QGDLIELHRHYIQDNKVIESAVVNISGPPKINFINDKYCAAT---GANEYMRLLGGTKQM 348

Qy 353 KKATSGGMVLVMSLWDDYYANMLWLDS TYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
| | |||| |::| | | | | | | : | | | : | | |
Db 349 GDAMSRGMVLAMSVWWSEGDFMAWL DQ-----GVAGPCDATEGDPKNIVKVQP 396

QY 413 NAKVTFSNIRKFGP1GSGTGNPSGGNPPG 439
| :|||||: | |||| : |||
Db 397 NPEVTFSNIRIGEIGGSTSSVKAPAYPG 423

RESULT 40

US-08-483-432-25

; Sequence 25, Application US/08483432

; Patent No. 5763254

Db 197 ---FINGVGNIK-----GQGVCCNELDIWEANSRATHIAPHPCSKPGLYGC 239
 Qy 239 EGDGCGGTYSNDRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTQFETSG 298
 || || : | | | | | | | | : | | | | | | | | | | | : || | : | | | : |
 Db 240 TGDECGSS-----GFCDKAGCGWNHNRINVTDFYGRGKQYKVDSTRKFTVTSQFVANK 292
 Qy 299 ----AINRYYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEEAEFGSSFSDKGGLTQF 352
 ::|:|:|:|:| : | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 293 QGDLIELHRHYIQDNKVIESAVVNISGPPKINFINDKYCAAT---GANEYMRLGGTKQM 348
 Qy 353 KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 349 GDAMSRGMVLAMSVWWSEGDFMAWLDQ-----GVAGPCDATEGDPKNIVKVQP 396
 Qy 413 NAKVTFSNIKFPIGSTGNPSGGNPPG 439
 | :| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
 Db 397 NPEVTFSNIRIGEIGSTSSVKAPAYPG 423

RESULT 41

US-08-709-974A-6

; Sequence 6, Application US/08709974A

; Patent No. 6117664

; GENERAL INFORMATION:

; APPLICANT: Sch lein, Martin

; APPLICANT: Rosholm, Peter

; APPLICANT: Nielsen, Jack Bech

; APPLICANT: Hansen, Svend Aage

; APPLICANT: von der Osten,Claus

; TITLE OF INVENTION: No. 6117664el Alkaline Cellulases

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6117664o No. 6117664disk of No. 6117664th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/709,974A

; FILING DATE: 09-SEP-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Gregg, Valeta

; REGISTRATION NUMBER: 35,127y

; REFERENCE/DOCKET NUMBER: 4160.414-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 411 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-709-974A-6

Query Match 27.1%; Score 741.5; DB 2; Length 411;

Best Local Similarity 38.0%; Pred. No. 4.6e-49;

Matches 170; Conservative 52; Mismatches 162; Indels 63; Gaps 15;

; GENERAL INFORMATION:

; APPLICANT: Woeldike, Helle F.
; APPLICANT: Hagen, Frederick
; APPLICANT: Hjort, Carsten M.
; APPLICANT: Sven, Hastrup

; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

; TITLE OF INVENTION: or Hemicellulose

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5457046o No. 5457046disk of No. 5457046th America, Inc.

; STREET: 405 Lexington Avenue, 62nd Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6201

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/361,920

; FILING DATE:

; CLASSIFICATION: 435

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/940,860

; FILING DATE: 28-OCT-1992

; APPLICATION NUMBER: DK 1158/90

; FILING DATE: 09-MAY-1990

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/DK91/00124

; FILING DATE: 08-MAY-1991

ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 3435.204-US

TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-867-0298

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:

; LENGTH: 435 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-361-920-27

Query Match 26.9%; Score 737.5; DB 1; Length 435;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 169; Conservative 59; Mismatches 155; Indels 69; Gaps 16;

Qy 9 ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATN--SSTNCYD-GNTWSSTLCVDNE 65
| | | | : | : | | | | : | : | | | | : | | | | : | | | |

Db 28 EVHPQLTTFRCTKRGCKPATNFIVLDSLSPHRAEGLGPGCGDWGNPPPKDVCVDNE 87

Qy 66 XCAKNCLDGAAYASTYGVTTSGNSLISGFVTQSAQNVGARLYLMASDTT--YQEFTL 122
: | | | | : | : | | | | | | | | | | | | | | | | | | | | | |

Db 88 SCAKNCIMEGIPDYSQYGVTNGTSLRLQHILPDG-RVPSPRVYLL--DKTKRYYEMHL 144

Qy 123 LGNEFSFDVDVSQPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC---PRD 179
| | | : | | | : | | | | | | | | | | | | | | | | | | | | | | | | |

Db 145 TGFEFTFDVDAKLPCGMNSALYLSEMHPTGAKSKY--NSGGAYYGTGYCDAQCFVTP-- 200

Qy 180 LKFINGQANVEGWEPSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
| | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 201 --FINGLGNIE-----GKGSCCNEMDIWEVNSRASHVVPHTCNKGLYLC 244

Qy 240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTFETS-- 297
| : | : | | : | | | | : | : | | | | | : | | | | | : | | | | | : |
Db 245 GEECA-----FEGVCDKNGCGWNYYRVNVTDYYGRGEFKVNTLKPFTVVTQFLANRR 297

Qy 298 ---GAINRYYVQNGVTFQQ--PNAELGSYSGNELNDDYCTAEEAEFGSSFSDKGLTQF 352
| : | : | | : | : | | | | : | : | : | : | | | | | : |
Db 298 GKLEKIHRFYVQDGKVIESFYTNKEGVPYT-NMIDDEFCEAT---GSRKYMELGATQGM 352

Qy 353 KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
| : | | | | : | | | | | | | | | | | | | | | | | | | : |
Db 353 GEALTRGMVLAMSIWWWDQGGNMELDH-----GEAGPCAKGEGAPSNIIVQVEP 400

Qy 413 NAKVTFSNIKFGPIGST-----GNPSGGNPP 438
| : | : | : | | | | | | | | | | | | | | | | | | | |
Db 401 FPEVTTYTNLRWGEIGSTYQEVKPKPKPGHGP 432

RESULT 45

US-08-479-939-27

; Sequence 27, Application US/08479939

; Patent No. 5686593

; GENERAL INFORMATION:

; APPLICANT: Woeldike, Helle F.

; APPLICANT: Hagen, Frederick

; APPLICANT: Hjort, Carsten M.

; APPLICANT: Sven, Hastrup

; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

; TITLE OF INVENTION: or Hemicellulose

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5686593o No. 5686593disk of No. 5686593th America, Inc.

; STREET: 405 Lexington Avenue, 62nd Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6201

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/479,939

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/361,920

; FILING DATE: 22-DEC-1994

; APPLICATION NUMBER: US 07/940,860

; FILING DATE: 28-OCT-1992

; APPLICATION NUMBER: DK 1158/90

; FILING DATE: 09-MAY-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/DK91/00124

; FILING DATE: 08-MAY-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 3435.204-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-867-0298

; INFORMATION FOR SEQ ID NO: 27:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 435 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein
US-08-479-939-27

Query Match 26.9%; Score 737.5; DB 1; Length 435;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 169; Conservative 59; Mismatches 155; Indels 69; Gaps 16;

Qy 9 ETHPPLTWQKCSSGGTCTQQTGSVIDANWRWTHATN--SSTNCYD-GNTWSSTLCPDNE 65
| | | | : | : | | | : | : | | | | : | | |
Db 28 EVHPQLTTFRCTKRGGCKPATNFIVLDSLSPHIHRAEGLPGGGCGDWGNPPPKDVCVD 87

Qy 66 XCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTT---YQEFTL 122
: | | | | : | | | | | : | : | : | | | | : |
Db 88 SCAKNCIMEGIPDYSQYGVTTNGTSRLQHILPDG-RVPSPRVYLL--DKTKRRYEMLHL 144

Qy 123 LGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC---PRD 179
| | | : | | | : | | | | | | | | | | | | : | |
Db 145 TGFEFTFDVDAKLPCGMNSALYLSEMHPTGAKSKY--NSGGAYYGTGYCDAQCFVTP-- 200

Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
| | | | | : | | | | | | | | | | | | | | | | : |
Db 201 --FINGLNIE-----GKGSCCNEMDIWEVNSRASHVVPHTCNKGLYLC 244

Qy 240 GDGCGGTYSNDNRYGGCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLTQFETS-- 297
| : | : | | : | | | | | : | | | | | | | | | | : |
Db 245 GEECA-----FEGVCDKNGCGWNNYRVNVTDYYGRGEFKVNTLKPFVVTQFLANRR 297

Qy 298 ---GAINRYVQNGVTFQQ--PNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGLTQF 352
| : | : | | : | : | | | | | | | | | | | | | | : |
Db 298 GKLEKIHRFYVQDGKVIESTYTNKEGPYT-NMIDDEFCEAT---GSRKYMELGATQGM 352

Qy 353 KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
| : | : | | | | : | | | | | | | | | | | | | | : |
Db 353 GEALTRGMVLAMSIWWWDQGGNMELDH-----GEAGPCAKGEGAPSNIQVEP 400

Qy 413 NAKVTFSNIKFGPIGST-----GNPSGGNPP 438
| : | : | : | | | | | | | | : |
Db 401 FPEVTYTNLRWGEIGSTYQEVQKPKPKPGHGP 432

RESULT 46

US-08-483-432-27

; Sequence 27, Application US/08483432

; Patent No. 5763254

; GENERAL INFORMATION:

; APPLICANT: Woeldike, Helle F.

; APPLICANT: Hagen, Frederick

; APPLICANT: Hjort, Carsten M.

; APPLICANT: Sven, Hastrup

; TITLE OF INVENTION: An Enzyme Capable of Degrading Cellulose

; TITLE OF INVENTION: or Hemicellulose

; NUMBER OF SEQUENCES: 85

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5763254o No. 5763254disk of No. 5763254th America, Inc.

; STREET: 405 Lexington Avenue, 62nd Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6201

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/483,432

; FILING DATE: 07-JUN-1995

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/361,920
FILING DATE:
APPLICATION NUMBER: US 07/940,860
FILING DATE: 28-OCT-1992
APPLICATION NUMBER: DK 1158/90
FILING DATE: 09-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00124
FILING DATE: 08-MAY-1991
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 3435.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-0298
FORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 435 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-483-432-27

Query Match 26.9%; Score 737.5; DB 1; Length 435;
Best Local Similarity 37.4%; Pred. No. 1e-48;
Matches 169; Conservative 59; Mismatches 155; Indels 69; Gaps 16;

Qy 9 ETHPLTQKCSSGGTCTQQTGSVVIDANWRWTHATN--SSTNCYD-GNTWSSTLCPDNE 65
| | | | : | : | | | : | : | | | | : | | |
Ph 28 EVHPLTTERCTKRGGCKPATNEVLDLSI.SHPTRHAFGLGPCCCGDWGNRRPKDVCVDVE 87

Qy 66 XCAKNCCLDGAAYASTYGVTTSGNLSIGFVTQSAQKNVGARLYLMAASDTT---YQEFTL 122
Db 88 SCAKNCIMEGIPDYSQYGVTNTGTSRLRQHILPDG-RVPSPRVYLL--DKTKRRYEMLHL 144

Qy 123 LGNEFSFDVDSLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC---PRD 179
| | | : | || | : : | || | : | | | | | | | | | | | | | | | | | | | | | | |
Db 145 TGFEFTFDVDAKLPCGMNSALYLSEMHPKGAKSKY--NSGGAYYGTGYCDAQCFVTP-- 200

Qy 180 LKFINGQANVEGWEPESSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
| | | | | : | | | | | : | | | | | : | | | | | : | |
Pb 201 --FINGLGNIE-----GKGSCCNEMDIWEVNSRASHVVPHTCNKKGLYLCE 244

Qy 298 ---GAINRYYVQNGVTFQQ--PNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLTQF 352
 | : | : | : | : | | | : | : | : : | : | : : |
 Db 298 GKLEKIHRFYVQDGKVIESFYTNKEGVPYT-NMIDDEFCEAT---GSRKYMELGATQGM 352

QY	413 NAKVTFSNIKFGPIGST-----GNPSSGGNPP 438
	: : :: :
Db	401 FPEVTYTNLRWGEIGSTYQEVKPKPKPGHGP 432

RESULT 47

US-09-069-632-3

; Sequence 3, Application US/09069632

; Patent No. 6261828

; MOLECULE TYPE: protein
US-08-709-974A-4

Query Match 26.8%; Score 735.5; DB 2; Length 415;
Best Local Similarity 37.4%; Pred. No. 1.4e-48;
Matches 169; Conservative 58; Mismatches 156; Indels 69; Gaps 16;

Qy 9 ETHPPLTWQKCSSGGTCTQQTGSVIDANWRWTHATN--SSTNCYD-GNTWSSTLCPDNE 65
| | | | | : | : | | | | : | : | | | | : | | | |
Db 8 EVHPQLTTFRCTKRGGCKPATNFIVLDSLSPHIHRAEGLPGGGCGDWGNPPPKDVCVDNE 67

Qy 66 XCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTT---YQEFTL 122
: | | | | | : | | | | | | | | | | | | | | | | |
Db 68 SCAKNCIMEGIPDYSQYGVTTNGTSRLQHILPDG-RVPSPRVYLL--DKTKRKYEMLHL 124

Qy 123 LGNEFSFDVDVSQPLCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC---PRD 179
| | | : | | | | : | | | | | | | | | | | | | | |
Db 125 TGFEFTFDVDAKLPGMNSALYLSHMPTGAKSKY--NPGGAYYGTGYCDAQCFVTP-- 180

Qy 180 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
| | | | | : | | | | | | | | | | | | | | | | | | : | |
Db 181 --FINGLNIE-----GKGSCCNEMDIWEVNSRASHVVPHTCNKGLYLCE 224

Qy 240 GDGCGGTYSDNRYYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTQFETS-- 297
| : | | | : | | | | | | | | | | | | | | | | | | | | | | : |
Db 225 GEECA-----FEGVCDKNGCGWNNYRVNVTDYYGRGEFKVNTLKPFVVTQFLANRR 277

Qy 298 ---GAINRYVQNGVTFQQ--PNAELGSYSGNELNDDYCTAEEAEFGSSFSDKGLTQF 352
| : | : | | | : | | | | | | | | | | | | | | | | | | | : |
Db 278 GKLEKIHRFYVQDGKVIESTYTNKEGPYT-NMIDDEFCEAT---GSRKYMELGATQGM 332

Qy 353 KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412
| : | : | | | | : | | | | | | | | | | | | | | | | | | | : |
Db 333 GEALTRGMVLAMSIWWDQGGNMELDH-----GEAGPCAKGEGAPSNIQVQEP 380

Qy 413 NAKVTFSNIKFGPIGST-----GNPSGGNPP 438
| : | : | : | | | | | | | | | | | | | | | | | | | | | | : |
Db 381 FPEVTTYTNLRWGEIGSTYQEVQKPKPKPGHGP 412

RESULT 50

US-08-709-979A-3

; Sequence 3, Application US/08709979A

; Patent No. 5912157

; GENERAL INFORMATION:

; APPLICANT: Claus von der Osten

; APPLICANT: Martin Schlein

; TITLE OF INVENTION: No. 5912157el Alkaline Cellulases

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5912157o No. 5912157disk of No. 5912157th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/709,979A

; FILING DATE: 09-SEP-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4160.404-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 402 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-709-979A-3

Query Match 26.5%; Score 725.5; DB 1; Length 402;
Best Local Similarity 37.7%; Pred. No. 7.7e-48;
Matches 166; Conservative 59; Mismatches 146; Indels 69; Gaps 16;

Qy 9 ETHPPLTWQKCSSGGTCTQQTGSVIDANWRWTHATNSST-----NCYD-GNTWSSTLCP 62
| | | | : | : | | | : | : : | : : | | || : ||
Ph 8 EIVVBDLTTERCTKPCCCCKPATNEIVY DLSLISVIRHAFEGI GPGCCCCDVKHRRRDKVGP 61

Qy 63 DNEXCAKNCLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNGARLYLMASDTT---YQE 119
| | : || || : | | || || : | | : : | : | : | | : | : | :
Ph 65 DVESCAKNCLMEGIPDYSOYGVTNTGTSIPLQHILPDG-PVPSPPVYLL-DKTKPRYEM 121

Qy 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQC-- 176

Ov 177 PRDLKFEFINGOANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGOE 236

Db 180 P----FINGLGNIE-----GKGSCCNEMDIWEANSRASHVAPHTCNKKGLY 221

Db 222 LCEGEECA-----FEGVCDKNGCGWNNYRVNVTDYGRGEFKVNTLKPFTVVTQFLA 274

Qy 297 S-----GAINRYYVQNGVTFQQ--PNAELGSYSGNELNDDYCTAEEAEFGSSFSDKGG 349
 : |:|:||:|:| : | | | : | : ||:| | | : |
 Db 275 NRRGKLEKIHRFYVQDGKVIESFYTNKEGVPYT-NMIDDEFCEAT---GSRKYMELGAT 329

Db 330 QGMGEALTRGMVLAMS IWWDQGGNMELDH-----GEAGPCAKGEGAPSNIQ 377

Qy 410 QSPNAKVTFSNIKFGPIGST 429
| :||::|:::| |||||
Pb 376 VEDERDEYVWVHPSVQASLALST 427

RESULT 51

REF ID: A643A-5

: Sequence 5 Application US/08833612A

, sequence 5, Applied
: Patent No. 5883066

: GENERAL INFORMATION:

APPLICANT: Ivan M. A. J. Herbots et al.

TITLE OF INVENTION: Liquid Detergent Compositions

TITLE OF INVENTION: Co

; NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

; ADDRESSEE: Jackie Ann Zurich

; ADDRESSEE: Dinsmore & Shoh

; STREET: 255 E. F

; STREET: 1900

COUNTRY: USA
ZIP: 45202
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/833,642A
FILING DATE: April 8, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Zurcher, J. A.
REGISTRATION NUMBER: P42,251
REFERENCE/DOCKET NUMBER: CM551C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (513) 977-8377
TELEFAX: (513) 977-8141
FORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 415 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-833-642A-5

Query Match 25.7%; Score 705.5; DB 1; Length 415;
Best Local Similarity 37.3%; Pred. No. 2.8e-46;
Matches 163; Conservative 59; Mismatches 152; Indels 63; Gaps 15;

QY 9 ETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATN--SSTNCYD-GNTWSSTLCPDNE 65
| | | | : | | | : | : | : | | | | : | | |

Qy 66 XCAKNCCLDGAAYASTYGVTTSGNLSIGFVTQSAQKNVGARLYLMASDTT---YQEFTL 122
Pj :||||| ::| | ||||:| :| :| :| :| :| :| :| :| :| :| :| :|

Qy 123 LGNEFSFDVDSLPCGLNGALYFVSMDAGGVSKYPTNTAGAKYGTGYCDSQC---PRD 179
 | ||::||| : :||| : | ||| | | | : | ||| ||||| : ||| | |
 Pb 125 TGFEEFTTERIDATKYLPGGMNCLVLYFENWDTGCKVSKVY NCCSCLVXSTGKPLGSCVYD 180

QY 180 LKFINGQANVEGWEPSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
 ||||| :| :| | | | :| | | | | | | :| :| | | | | :| |
 Ph 181 --EINCLGNIE-- GKGSGCCNEMDIWEVNSRASVHIVPTCNKKGLYCE 234

Qy 240 GDGCGGTYSNDRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDDTKKLTVVTQFETS-- 297

QY 298 ---GAINRYYVQNGVTFQQ--PNAELGSYSGNELNDDYCTAEEAEFGGSSFSDKGGLTQF 352
| : | : | | : | : | | | | : | : | : | : | | | : | | |

Qy 353 KKATSGGMVLVMSLWDDYYANMLWLSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 412

Ov 413 NAKVTFESNIKFGPIGST 429

Db 381 FPEVTVTNLRWGEIGST 397

RESULT 4
S38784

338734
selluk

C;Species: *Humicola grisea* var. *thermoidea*

Qy	415 KVTFSNIKFGPI-----GSTGNPSGGNPPGGNNPGTTTRRPATTTGSSP-GPTQSHY 466 : :: :: : :: :: : :
Db	427 KVAFSNIKFGHLGISPFSGGSSGTTPP-SNPSSSASPSTSSTAKPSSTSTASNPSGTGAHW 485
Qy	467 GQC GGIGYSGPTVCASGTTCQVLNPYYSQCL 497 : : :
Db	486 AQCGGIGFSGPTTCPEPYTCAKDHDYISQCV 516

RESULT 13

JE0313

exoglucanase (EC 3.2.---) - imperfect fungus (*Humicola grisea*)

C; Species: *Humicola grisea*

C;Date: 05-Feb-1999 #sequence revision 05-Feb-1999 #text change 09-Jul-2004

C;Accession: JE0313

R;Takashima, S.; Iikura, H.; Nakamura, A.; Hidaka, M.; Masaki, H.; Uozumi, T.
J. Biochem. 124, 717-725, 1998

A;Title: Isolation of the gene and characterization of exoglucanase of *Humicola grisea* without a cellulose-bin

A: Reference number:

A:Reference Number: 8
A:Accession: JE0313

A:Accession: 000313
A:Status: preliminary

A, Status: preliminary
A, Molecule type: RNA

A:Residues: 1-451 <TAK>

A:Cross-references: UNIPROT:Q93780; UNIPARC:UPI000005E865; DDDBI:AB003105

C;Superfamily: cellulose 1,4-beta-
C;Keywords: glycosidase; hydrolase

Query Match 45.4%; Score 1243.5; DB 2; Length 451;
Best Local Similarity 52.0%; Pred. No. 3.3e-67;
Matches 226; Conservative 85; Mismatches 113; Indels 11; Gaps 9;

Qy 61 CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQ-KNVGARLYLMASDTTYQE 119
| :||: | |||| | ||| :|||:||:: |||: | :|||: ||||: ||
Pb 80 CSSATDCAORCALDGANYOSTYGAESTGDSLTLKFEVTKHEVGTNIGSREYLMANONKYOM 139

Qy 120 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
|||: |||: |||: |:: |||: |||||: |: |||: |||: |||: |||: |||: |||: |||: |||:
Db 140 FTLMNNEFAFDVDSLKVECGINSALYFVAMEEDGGMASYPSNRAGAKYGTGYCDAOCARD 199

Qy 180 LKFINGQANVEGWEPSNNANTGIGGHGCCSEMDIWEANSISEALTPHPCTTVGQ-EIC 238
 ||||| : | : ||| : ||| : | : | : | : | : | : | : | : | : | : | : |
 Db 200 LKFIGGKANIEWRPSTNDPNAGVGPAGACCAEIDVWESNAYAYAFTPAGCSKNRYHIC 259

Qy 239 EGDGCGGTYSNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLTVVTFETSG 298
| : ||||| : | : | | : ||| : ||| : | | | | | : | : | | | : | : | : | : | : |
Db 260 ETNNCGGTYSDDRFAGYCDANGCDYNPYRMGNKDFYGKGK--TVDTNRKFTVVSRFERN- 316

Qy 299 AINRYVQNGVTQQPNAEL-GSYSGNELNDDYCTAEEAEFGG-SSFSDKGGLTQFKKAT 356
Db : :::::||:| : | : | : : | |: | : | : || : | : |
317 RLSQFFVQDGRKIEVPPPTWPGLPNSADITPELCDAQFRVFDDNRFAETGGFDALNEAL 376

Qy 357 SGGMVLVMSLWDDYYANMLWLDSYPTINETSSTPGAVRGSCSTSSGVPAQVESQSPNAKV 416
 : |||||:|||::|||||:|| :|| :|| |:|||:|||:|||:
Db 377 TIPMVLVMSIWDDHHSNMLWLDSYPP-EKAGLPGGDRGCPCTTSGVPAEVEAQYPDAQV 435

QY 417 VWSNIRFGPIGSTVN 450
Db :|||||:|||||||

Q12621_HUMGT
ID Q12621_HUMGT PRELIMINARY; PRT; 525 AA.
AC Q12621;
DT 01-NOV-1996, integrated into UniProtKB/TrEMBL.
DT 01-NOV-1996, sequence version 1.
DT 07-FEB-2006, entry version 31.
DE Cellulase (EC 3.2.1.91).
GN Name=cbh-1;
OS Humicola grisea var. thermoidea.
OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
OX NCBI_TaxID=5528;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IFO9854;
RA Takashima S., Nakamura A., Hidaka M., Masaki H., Uozumi T.;
RT "Cloning, sequencing, and expression of the cellulase genes of
RT Humicola grisea var. thermoidea.";
RL Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: The biological conversion of cellulose to glucose
CC generally requires three types of hydrolytic enzymes: (1)
CC Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
CC Exocellobiohydrolases that cut the disaccharide cellobiose from
CC the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-
CC glucosidases which hydrolyze the cellobiose and other short cello-
CC oligosaccharides to glucose (By similarity).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
CC in cellulose and cellotetraose, releasing cellobiose from the non-
CC reducing ends of the chains.
CC -----
CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
CC Distributed under the Creative Commons Attribution-NoDerivs License
CC -----
DR EMBL; D63515; BAA09785.1; -; Genomic_DNA.
DR HSSP; Q09431; 1GPI.
DR GO; GO:0005576; C:extracellular region; IEA.
DR GO; GO:0016162; F:cellulose 1,4-beta-celllobiosidase activity; IEA.
DR GO; GO:0030248; F:cellulose binding; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030245; P:cellulose catabolism; IEA.
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.
DR InterPro; IPR000254; CBD_fun.
DR InterPro; IPR001722; Glyco_hydro_7.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF00840; Glyco_hydro_7; 1.
DR PRINTS; PR00734; GLHYDRLASE7.
DR ProDom; PD001821; CBD_fungal; 1.
DR ProDom; PD186135; Glyco_hydro_7; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
KW Carbohydrate metabolism; Cellulose degradation; Glycosidase;
KW Hydrolase; Polysaccharide degradation.
SQ SEQUENCE 525 AA; 55722 MW; A2E6E5F40F6D3BB0 CRC64:

Query Match 60.5%; Score 1658; DB 2; Length 525;
Best Local Similarity 57.5%; Pred. No. 5.9e-101;
Matches 295; Conservative 77; Mismatches 119; Indels 22; Gaps 7;

Qy 61 CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA-QKNGARLYLMASDFTTYQE 119
| | : || : || | : || | : || | : || : || | : || | : || | : || | : ||
Pb 79 CTDAKSCAONCCVDGADYTSTYGITNGDSLKLKEVTKGOYSTNVGSRTYLMDGEDKYQT 138

Qy 120 FTLLGNEFSFDVDVSQPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRD 179
| :||||||:||||| :|||||||||:||||:|||:|||||:|||||

Db 139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMADGLSRYPGNKAGAKYGTGYCDAQCPRD 198
 Qy 180 LKFINGQANVEGWEPPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 239
 :|||||:|||:||| |:||| | |:||| |:||| |:||| |:|||
 Db 199 IKFINGEANIEGWTGSTNDPNAGAGRGTCCSEMDIWEANNMATAFTPHTPCTIIGQSRCE 258
 Qy 240 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTFETS-- 297
 | | | | |: | | | | | | | | | | |: | |
 Db 259 GDSCGGTYSNERYAGVCDPDGCFNSYRQGNKTFYKGK--MTVDTTKITVVTFQLKDAN 316
 Qy 298 ---GAINRYYQNGVTQQPNAELGSYSGNELNDDYCTAEEAEFGG-SSFSDKGLTQFK 353
 | | | | |: | | | | | | | | | | |: | |
 Db 317 GDLGEIKRFYVQDGKIIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG 376
 Qy 354 KATSGGMVLVMSIWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 413
 | | | | |: | | | | | | | | | | |: | |
 Db 377 KALAGPMVLVMSIWDDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAPN 435
 Qy 414 AKVTFSNIKFGPIGST-----GNPSGGNPPGNPPGTTTRPATTGSSPGPTQS 464
 : | | | |: | | | | | | | | | | | | | |
 Db 436 SNVVFSNIRFGPIGSTVAGLPAGNGGNNGNPP---PPTTTSSAPTTTASAGPKAG 492
 Qy 465 HYGQC GGIGYGSGPTVCASGTTCQVLNPYYSQCL 497
 : | | | |: | | | | | | | | |: | |
 Db 493 RWQQC GGI GFTGPTQCEEPYTCTKLNDWYSQCL 525

RESULT 15

GUX1_HUMGT

ID GUX1_HUMGT STANDARD; PRT; 525 AA.
 AC P15828;
 DT 01-APR-1990, integrated into UniProtKB/Swiss-Prot.
 DT 01-FEB-1996, sequence version 2.
 DT 07-FEB-2006, entry version 55.
 DE Exoglucanase 1 precursor (EC 3.2.1.91) (Exoglucanase I)
 DE (Exocellobiohydrolase I) (1,4-beta-cellulobiohydrolase) (Beta-
 DE glucancellulobiohydrolase).
 GN Name=CBH-1;
 OS Humicola grisea var. thermoidea.
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 OX NCBI_TaxID=5528;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
 RX MEDLINE=90175006; PubMed=2308855;
 RA de Oliveira Alzevedo M., Radford A.;
 RT "Sequence of cbh-1 gene of Humicola grisea var. thermoidea.";
 RL Nucleic Acids Res. 18:668-668(1990).
 CC -!- FUNCTION: The biological conversion of cellulose to glucose
 generally requires three types of hydrolytic enzymes: (1)
 Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2)
 Exocellobiohydrolases that cut the disaccharide cellobiose from
 the nonreducing end of the cellulose polymer chain; (3) Beta-1,4-
 glucosidases which hydrolyze the cellobiose and other short cello-
 oligosaccharides to glucose.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 in cellulose and cellobotetraose, releasing cellobiose from the non-
 reducing ends of the chains.
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 7 (cellulase C)
 family.
 CC -!- SIMILARITY: Contains 1 CBM1 (fungal-type carbohydrate-binding)
 domain.
 CC -----
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>
 CC Distributed under the Creative Commons Attribution-NoDerivs License
 CC -----
 DR EMBL; X17258; CAA35159.1; -; Genomic_DNA.

DR PIR; S38794; S38794.
 DR HSSP; Q09431; 1GPI.
 DR InterPro; IPR000254; CBD_fun.
 DR InterPro; IPR001722; Glyco_hydro_7.
 DR Pfam; PF00734; CBM_1; 1.
 DR Pfam; PF00840; Glyco_hydro_7; 1.
 DR PRINTS; PR00734; GLHYDRASE7.
 DR ProDom; PD001821; CBD_fungal; 1.
 DR ProDom; PD186135; Glyco_hydro_7; 1.
 DR SMART; SM00236; fCBD; 1.
 DR PROSITE; PS00562; CBM1_1; 1.
 DR PROSITE; PS51164; CBM1_2; 1.
 KW Carbohydrate metabolism; Cellulose degradation; Glycoprotein;
 KW Glycosidase; Hydrolase; Polysaccharide degradation; Signal.
 FT SIGNAL 1 18 Potential.
 FT CHAIN 19 525 Exoglucanase 1.
 FT /FTId=PRO_0000007921.
 FT DOMAIN 489 525 CBM1.
 FT REGION 19 467 Catalytic.
 FT REGION 468 489 Linker.
 FT ACT_SITE 231 231 Nucleophile (By similarity).
 FT ACT_SITE 236 236 Proton donor (By similarity).
 FT CARBOHYD 289 289 N-linked (GlcNAc. . .) (Potential).
 FT DISULFID 497 514 By similarity.
 FT DISULFID 508 524 By similarity.
 SQ SEQUENCE 525 AA; 55694 MW; A6684D4CF881E090 CRC64;

Query Match 60.3%; Score 1652; DB 1; Length 525;
 Best Local Similarity 57.3%; Pred. No. 1.5e-100;
 Matches 294; Conservative 76; Mismatches 121; Indels 22; Gaps 7;

Qy	1	QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL	60
Db	19	QQACSLTTERHPSLSWNKCTAGQCQTVQASITLDSNWRTHQVSGSTNCYTGNKWDTI	78
Qy	61	CPDNEXCAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQK-NVGARLYLMASDTTYQE	119
Db	79	CTDAKSCAQNCCVDGADYTSTYGITNGDSLSSLKFVTKQHSTNVGSRTYLMGEDKYQT	138
Qy	120	FTLLGNEFSFDVDSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRD	179
Db	139	FELLGNEFTFDVDSNIGCGLNGALYFVSMADGGLSRYPGNKAGAKYGTGYCDAQCPRD	198
Qy	180	LKFINGQANVEGWEPESSNNANTGIGGHGSCCSEMDIWEANSISEALT PHPCTTVGQEICE	239
Db	199	IKFINGEANIEGWTGSTNDPNAGAGRYGTCCSEMDIWEANNMATAFTPHTCIIIGQSRCE	258
Qy	240	GDGCCGTYSNDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKKLTVVTQFETS--	297
Db	259	GDSCGGTYSNERYAGVCDPDGCDFNRSYRQGNKTFYGKG--MTVDTTKITVVTQFLKDAN	316
Qy	298	--GAINRYVQNGVTFQQPNAELGSYSGNELNDYCTAEEAEFGG-SSFSDKGLTQFK	353
Db	317	GDLGEIKRFYVQDGKIIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG	376
Qy	354	KATSGGMVLVMSLWDDYYANMLWLDSYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN	413
Db	377	KALAGPMVLVMSIWDDHASNMLWLDSFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN	435
Qy	414	AKVTFSNIKFGPIGST-----GNPSGGNPPGNPPGTTTRPATTGSSPGPTQS	464
Db	436	SNVVFNSIRFGPIGSTVAGLPAGNGGNNGGNPP--PPTTTSSAPATTTASAGPKAG	492
Qy	465	HYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL	497
Db	493	RWQQCGGIGFTGPTQCEEPYICTKLNDWYSQCL	525

Db 146 FTFDVDVKVPCGINGALYFIQMDADGGMSKQPNNRAGAKYGTGYCDSQCPRDIKFIDGV 205
 Qy 186 ANVEGWEPPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTV---GQEICEGDGC 242
 || | || : | | : | ||:|||:||| | ||| | | : || |
 Db 206 ANSADWTPSETDPNAGRGRYRICCAEMDIWEANSISNAYTPHPCRTQNDGGYQRCEGRDC 265
 Qy 243 GGTYSNDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTFET----SG 297
 : | | |||||:||:|| | ||| | :|| :||:||| | :|
 Db 266 ---NQPRYEGLCDPDGCDYNPFRMGNKDFYGPBK--TVDTNRKMTVVTQFITHDNTDTG 319
 Qy 298 A---INRYYYQNGVTFQQPNAELGSY--SGNELNDDYCTAEEAEFGG-SSFSDKGLTQF 351
 | | |||: | | : : : :|| : || | ||: | |||:
 Db 320 TLVDIRRLYVQDGRVIANPPTNFPGLMPAHDSITEQFCTDQKNLFGDYSSFARDGGLAHM 379
 Qy 352 KKATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSP 411
 :: : | || :||:||: | :||| | ||: : | || ||: | : | | : |
 Db 380 GRSLAKGHVLALSIWNDHGAHMLWLDSNYPTADPNKPGIARGTCPTGGTPRETEQNHP 439
 Qy 412 NAKVTFSNIKFGPIGST 428
 :||: | ||||| | |||
 Db 440 DAQVIFSNIKFGDIGST 456

RESULT 171

AAR94351

ID AAR94351 standard; protein; 451 AA.

XX

AC AAR94351;

XX

DT 29-AUG-1996 (first entry)

XX

DE Humicola insolens cellulase.

XX

KW Cellulase; detergents; textile auxiliaries; feed additives;

KW digestive agents; host cell; recombinant production.

XX

OS Humicola insolens.

XX

FH Key Location/Qualifiers

FT Peptide 1. .22

FT /label= sig_peptide

FT Peptide 23. .451

FT /label= mat_peptide

XX

PN JP08056663-A.

XX

PD 05-MAR-1996.

XX

PF 29-AUG-1994; 94JP-00203564.

XX

PR 29-AUG-1994; 94JP-00203564.

XX

PA (MEIJ) MEIJI SEIKA KAISHA LTD.

XX

DR WPI; 1996-182296/19.

DR N-PSDB; AAT13426.

XX

PT Humicola insolens cellulase - used as main component in detergents,
 PT textile auxiliaries, feed additives and digestive agents.

XX

PS Claim 2; Page 9-10; 16pp; Japanese.

XX

CC The present sequence is H. insolens cellulase, which is used as the main
 CC component in detergents, textile auxiliaries, feed additives and
 CC digestive agents. A host cell transformed with a vector contg. the
 CC cellulase DNA, can be used for the recombinant prodn. of the cellulase

XX

PR 24-JUL-1996; 96JP-00195070.

PA (MEIJI) MEIJI SEIKA KAISHA LTD

PI Moriya T, Murashima K, Aoyagi K, Sumida N, Watanabe M, Hamaya T;
PI Koqa J, Kono T, Murakami T;

DB WPI: 1998-120786/11

DR N=PSDB: AAV19376

PT Mass production of proteins and peptides in *Humicola* species - using
PT expression vector containing the promoter, signal sequence and/or
PT terminator from the *Humicola insolens* NCE1 or NCE2 gene.

PS Claim 8; Page 34-39; 63pp; Japanese.

The present sequence represents the Humicola insolens cellulase NCE1 protein from the present invention. The present invention describes a method for the mass production of proteins and peptides in Humicola species, especially in Humicola insolens, using an expression vector which comprises the promoter, signal sequence and/or terminator regulatory sequences from the NCE1 or NCE2 gene of H. insolens. These are available in the plasmids pM3-1 (Escherichia coli JM109/pM3-1, FERM BP-5971) (for NCE1) and pM14-1 (E. coli JM109/pM14-1, FERM BP-5972) (for NCE2). The vector also contains a marker gene such as an antibiotic resistance gene (e.g. the destomycin resistance gene from Streptomyces rimofaciens). Proteins which can be expressed using this system include cellulase, amylase, lipase, protease, phytase and other enzymes. Specific expression vectors of the invention are pMKD01 (for Humicola NCE3 cellulase gene), pEGD01 (for Humicola NCE4 cellulase gene) and pIED02 (for Humicola NCE4 cellulase gene). The expression system allows the efficient production of proteins and peptides in a Humicola host. Using the expression system high amounts of protein (>4.5 g/l) can be obtained

SQ Sequence 451 AA;

Query Match 45.6%; Score 1250; DB 2; Length 451;
Best Local Similarity 52.1%; Pred. No. 7e-73;
Matches 226; Conservative 84; Mismatches 114; Indels 10; Gaps 8;

Qy 61 CPDNECAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQ-KNVGARLYLMASDTTYQEF 119
::|: | ||| | | | | :|||:|||: | | | : | ; | | | | | | |

Qy 120 TLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVSKYPTNTAGAKYGTGYCDSQCPRDL 179
||: |||: ||||: :: ||: |||||: |: |||: : ||: |||||||: ||| |||
Ph 141 TLMLNEDRERDVLGIVESSGNGALVWVHMEEDSGVHGSQHLSLQHSHSPLAQSRS 200

Qy 180 KFINGQANVEGWEPPSNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQ-EICE 238
 ||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
 Ph 201 KEIGGGKANIEGWRPSTNDRNAGVYGRMGACGGAELDIVEGNAVAXAETDYNCCCKNRYVHICE 262

Qy 239 GDGCGGTYSNDRYGGTCDPDGCWDNPYRLGNTSFYGPSSFTLDTKKLTVVTFETSGA 298
 : ||||||| : | | : ||| : ||| : || | | | | | : | : | | | : | : | : | : |
Pb 261 TNNCGGTYSDDPEAGYCDANCCDYNRVRMCNKDEYCKCK TVDTNPKETVYVSREEDN P 317

Qy 299 INRYYYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAAEAEFGG-SSFSDKGGLTQFKKATS 356
 :::|:||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
 Db 318 LSOFFVODGRKTEVPPPTWPGI.PNSADITPEL.CDAQFRVEDDENPRAETGGEDAL.NEALT 377

Qy 357 GGMVLVMSLWDDYYANMLWLDSITYPTNETSSTPGAVRGSCSTSSGVPQAQVESQSPNAKVT 416
 |||||||:||||::|||||||:||| : ||| ||| |:|||||:|||:|||:
 Db 378 IIPMVLVMSIWDDHHSNMLWLDSSYPP-EKAGLPGGDRGPGCPTTSGVPAEVFAQVNPNAQVV 436

Qy 120 TLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRDL 179
 ||: |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 Db 141 TLMNNNEFAFDVDSLKVECGINSALYFVAMEEDGGMASYPSNRAGAKYGTGYCDAQCARDL 200

 Qy 180 KFINGQANVEGWEPSSNNANTGIGGHGCCSEMDIWEANSISEALTPHPCTTVGQ-EICE 238
 ||| :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 Db 201 KFIGGGKANIEGWRPSTNDPNAGVGPAGACCAEIDVWESNAYAYAFTPACGSKNRYHICE 260

 Qy 239 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTTKLTVVTFETSGA 298
 :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
 Db 261 TNNCGGTYSDDRFAGYCDANGCDYNPYRGMNKDFYGKGK-TVDTNRKFTVVSFERN-R 317

 Qy 299 INRYVQNGVTFOQPNAEL-GSYSGNELNDDYCTAEEAEFGG-SSFSDKGLTQFKKATS 356
 :::::|||:|||:|||:|||:|||:
 Db 318 LSQFFVQDGRKIEVPPPTWPGLPNSADITPELCDAQFRVFDDRNRAETGGFDALNEALT 377

 Qy 357 GGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVT 416
 |||:|||:|||:|||:|||:
 Db 378 IPMVLVMSIWDDHHSNMLWLDSSYPP-EKAGLPGGDRGPCPTTSGVPAEVEAQYPDAQVV 436

 Qy 417 FSNIKFGPIGSTGN 430
 :|||:|||:
 Db 437 WSNIRFGPIGSTVN 450

RESULT 9

US-09-463-712C-10

; Sequence 10, Application US/09463712C

; Patent No. 6558937

; GENERAL INFORMATION:

; APPLICANT: DSM, N.V.

; APPLICANT: Gielkens, Marcus

; APPLICANT: Vesser, Jacob

; APPLICANT: De Graaff, Leendert

; TITLE OF INVENTION: CELLULOSE DEGRADING ENZYMES OF

; TITLE OF INVENTION: ASPERGILLUS

; FILE REFERENCE: 24615-20135.00

; CURRENT APPLICATION NUMBER: US/09/463,712C

; CURRENT FILING DATE: 2000-04-04

; PRIOR APPLICATION NUMBER: PCT/EP98/05047

; PRIOR FILING DATE: 1998-07-31

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 10

; LENGTH: 536

; TYPE: PRT

; ORGANISM: Aspergillus niger

US-09-463-712C-10

Query Match 61.1%; Score 1673.5; DB 2; Length 536;
 Best Local Similarity 59.6%; Pred. No. 2.3e-122;
 Matches 308; Conservative 61; Mismatches 125; Indels 23; Gaps 7;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL 60
 | | :|||:|||:|||:|||:
 Db 22 QQVGTYTTEHPSLTWQTCTSDGSCTNDGEVVIDANWRWVHSTSSATNCYTGNEWDTI 81

Qy 61 CPDN-ECAKNCLLDGAAYASTYGVTTSGNSLSIGFVTQSAQKNVGARLYLMASDTTYQEF 119
 | | :|||:|||:|||:|||:
 Db 82 CTDDVTCAANCALDGATYEATYGVTTSGSELRLNFVTQGSSKNIGSRLYMSDDSNYELF 141

Qy 120 TLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRDL 179
 |||:|||:|||:|||:|||:
 Db 142 KLLGQEFTFDVDVSNLPCGLNGALYFVAMDADGGTSEYSGNKAGAKYGTGYCDSQCPRDL 201

Qy 180 KFINGQANVEGWEPSSNNANTGIGGHGCCSEMDIWEANSISEALTPHPCTTVGQEICEG 239

; . ORGANISM: H. grisea
US-08-676-166A-3

Query Match 60.1%; Score 1647.5; DB 1; Length 525;
Best Local Similarity 57.5%; Pred. No. 2.4e-120;
Matches 295; Conservative 75; Mismatches 120; Indels 23; Gaps 8;

Qy 1 QSACTLQSETHPPLTWQKCSSGGTCTQQTGSVVIDANWRWTHATNSSTNCYDGNTWSSTL 60
| ||::| :| || |:||:||:|| | | : :||:||||| : ||||| || | :::
Db 19 QQACSLTTERHPSLSWKCTAGGQCQTVQASITLDSNWRWTHQVSGSTNCYTGNKWDTSI 78

Qy 61 CPD-NECAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSA-QKNVGARLYLMASDTTYQE 118
| | | ||:||:|| | | ||:||:||:||:|| | | : | ||:| || | : ||
Db 79 CTDAKSCAQNCVCDGADYTSTYGITNGDSLKFVTKQYSTNVGSRTYLMGEDKYQT 138

Qy 119 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMDADGGVKYPTNTAGAKYGTGYCDSQCPRD 178
| | |||||:||| | : ||| | ||| | ||| | ||| | :||| | | ||| | :||| |
Db 139 FELLGNEFTFDVDVSNIGCGLNGALYFVSMDADGLSRYPGNKAGAKYGTGYCDAQCPRD 198

Qy 179 LKFINGQANVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALT PHPCTTVGQEICE' 238
: ||| | : ||| | | : | | | : | ||| | ||| | : | | | | : ||| |
Db 199 IKFINGEANIEGTGSTNDPNAGAGRYGTCCSEMDIWEANNMATAFTP PHPCTIIGQSRCE 258

Qy 239 GDGCGGTYSDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTFETS-- 296
| | | ||| | : | | | ||| | : | | | | : | | | | : | | | |
Db 259 GDSCGGTYSNERYAGVCDPDGCFNSYRQGNKTFYKGK--MTVDTTKITVVTFQLKDAN 316

Qy 297 ---GAINRYVQNGVTQQPNAELGSYSGNELNDDYCTAEAEFGG-SSFSDKGGLTQFK 352
| | | : ||| | : | : | | : | | | | : | | | | : | | | : |
Db 317 GDLGEIKRFYVQDGKIIIPNSESTIPGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMKQMG 376

Qy 353 KATSGGMVLVMSLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPN 412
| | : | | | | : | | : | | | | : | | | | : | | | | : | | |
Db 377 KALAGPMVLVMSIWDDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN 435

Qy 413 AKVTFSNIKFGPIGST-----GNPSGGNPPGGNPPGTTTRPATTGSSPGPTQS 463
: | | | | : | | | | | : | | | | | : | | | | : | | | |
Db 436 SNVVFSNIRFGPIGSTVAGLPGAGNGGNNGNPP---PPTTTSSAPATTTASAGPKAG 492

Qy 464 HYGQCGGIGYSGPTVCASGTTCQVLNPYYSQCL 496
: | | | | : | | | | | | | | : | | | |
Db 493 RWQQCGGIGFTGPTQCEEPYTCTKLNDWYSQCL 525

RESULT 19

US-09-329-350-35

; Sequence 35, Application US/09329350

; Patent No. 6184019

; GENERAL INFORMATION:

; APPLICANT: Miettinen-Oinonen, Arja
; APPLICANT: Londesborough, John
; APPLICANT: Vehmaanper , Jari
; APPLICANT: Haakana, Heli
; APPLICANT: M ntyl , Arja
; APPLICANT: Lantto, Raija
; APPLICANT: Elovainio, Minna
; APPLICANT: Joutsjoki, Vesa
; APPLICANT: Paloheimo, Marja
; APPLICANT: Suominen, Pirkko

; TITLE OF INVENTION: NOVEL CELLULASES, THE GENES ENCODING THEM AND

; TITLE OF INVENTION: USES THEREOF

; NUMBER OF SEQUENCES: 45

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.

Qy 187 NVEGWEPSSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEGDGCGGY 246
 |:|||: |::| :|||:||:|||: :| ||| ||| :||| |||||
 Db 210 NIEGWKSSTSDPNAGVGPYGSACAEIDWESNAYAFAFTPHACTTNEYHVCETTNCGGTY 269
 |:|||: |::| :|||:||:|||: ||| ||| :||| ||||| :||| :|||:
 Qy 247 SDNRYYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTQFETSGAINRYYVQN 306
 |:|||: |::| :|||:||:|||: ||| ||| ||| :||| :||| :||| :|||:
 Db 270 SEDRFAGKCDANGCDYNPYRMGNPDFYKGK--TLDTSRKFTVVSERFE-ENKLSQYFIQD 326
 |:|||: |::| :|||:||:|||: ||| ||| ||| :||| :||| :||| :|||:
 Qy 307 G--VTFQQPNAELGSYSGNELNDDYCTAEEAEFGG-SSFSDKGLTQFKKATSGGMVLVM 363
 |:|||: |::| :|||:||:|||: ||| ||| ||| :||| :||| :||| :|||:
 Db 327 GRKIEIPPPPTWE-GMPNSSEITPELCSTMFDVFNDRNRFEVGGEQLNNALRPMVLVM 385
 |:|||: |::| :|||:||:|||: ||| ||| ||| :||| :||| :||| :|||:
 Qy 364 SLWDDYYANMLWLDSTYPTNETSSTPGAVRGSCSTSSGVPAQVESQSPNAKVTFSNIKFG 423
 |:|||: |::| :|||:||:|||: ||| ||| ||| :||| :||| :||| :|||:
 Db 386 SIWDDHYANMLWLDSIYPP-EKEGQPGAARGDCPTDSGVPAEVEAQFPDAQVVWSNIRFG 444
 |:|||: |::| :|||:||:|||: ||| ||| ||| :||| :||| :||| :|||:
 Qy 424 PIGST 428
 |||||
 Db 445 PIGST 449

RESULT 33

US-08-709-974A-11

; Sequence 11, Application US/08709974A

; Patent No. 6117664

; GENERAL INFORMATION:

; APPLICANT: Sch lein, Martin

; APPLICANT: Rosholm, Peter

; APPLICANT: Nielsen, Jack Bech

; APPLICANT: Hansen, Svend Aage

; APPLICANT: von der Osten, Claus

; TITLE OF INVENTION: No. 6117664el Alkaline Cellulases

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6117664o No. 6117664disk of No. 6117664th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/709,974A

; FILING DATE: 09-SEP-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Gregg, Valeta

; REGISTRATION NUMBER: 35,127y

; REFERENCE/DOCKET NUMBER: 4160.414-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 456 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-709-974A-11

Query Match

27.4%; Score 751; DB 2; Length 456;

. Best Local Similarity 36.3%; Pred. No. 1.9e-50;
Matches 172; Conservative 71; Mismatches 161; Indels 70; Gaps 17;

Qy 9 ETPPLTWQKCSSGTCTQQTGSVVIDANWRWTHATNSSTNCYD-GNTWSSTLCPDNE-C 66
| || : | : | : | : | : | : | : | : | : | : | : |
Db 28 EVHPQITTYRCTKADGCEEKTNYIVLDALSHPVHQVDNPYNCGDWGQKPNETACPDLESC 87

Qy 67 AKNCLLDGAAYASTYGVTTSGNSLSIGFVTQSAQKN-VGARLYLMASDTT---YQEFTLL 122
| : | : | : | : | : | : | : | : | : | : | : |
Db 88 ARNCIMDPVSDYGRHGSTDGTSRL---KQLVGGNVSPRVYLL--DETKERYEMLKLT 142

Qy 123 GNEFSFDVDVSQQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQC---PRDL 179
| : | : | : | : | : | : | : | : | : | : | : |
Db 143 GNEFTFDVDAKLPCGMNSALYLSMEDIATGARSE--LNPGGATFGTGYCDAQCYVTP--- 197
| : | : |
Qy 180 KFINGQANVEGWEPSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICEG 239
| : | : | : | : | : | : | : | : | : | : |
Db 198 -FINGLGNIE-----GKGACCNEMDIWEANARAQHIAAPHPCSKAGPYLCEG 242

Qy 240 DGCGGTYSNDNRYGGTCDPDGCDWNPYRLGNTSFYGPSSFTLDTKKLTVVTQFETSG-- 297
| : | : | : | : | : | : | : | : | : | : | : |
Db 243 AEC-----EFDGVCDFNGCAWNPYRVNVTDYEGAEFRVDTTRPFSVVTQFRAGGDA 295

Qy 298 -----AINRYYVQNGVTFQQPNAEL-GSYSGNELNDDYCTAEEAEFGGSSFSDKGLTQ 350
: | : | : | : | : | : | : | : | : | : | : |
Db 296 GGGKLESIYRLFVQDGRVIESYVVDKPGPPTDRMTDEFCAAT---GAARFTELGAMEA 351

Qy 351 FKKATSGGMVLVMSLWDDYYANMLWLDSITYPTNETSSTPGAVRGSCSTSSGVPAQVESQS 410
| : | : | : | : | : | : | : | : | : | : |
Db 352 MGDALTRGMVLALSIWWSEGNDMNWLDS-----GEAGPCDPDEGNPSNIIRVQ 399

Qy 411 PNAKVTFSNIKFGPIGSTGNPSGGNPPGGNPPGTTTRPATTGSSPGPTQSH 464
| : | : | : | : | : | : | : | : | : | : |
Db 400 PDPEVVFSNLRWGEIGST-YESAVDGPVGKGKGKGKAPA---GDGNGKEKSN 449

RESULT 4

S38794

cellulose 1,4-beta-cellobiosidase (EC 3.2.1.91) - imperfect fungus (*Humicola grisea*)

N;Alternate names: beta-glucan cellobiohydrolase; exoglucanase

C;Species: *Humicola grisea* var. *thermoidea*

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C;Accession: S38794; S08240; A45869

R;Radford, A.

submitted to the EMBL Data Library, June 1991

A;Reference number: S38794

A;Accession: S38794

A;Molecule type: DNA

A;Residues: 1-525 <RAD>

A;Cross-references: UNIPROT:P15828; UNIPARC:UPI000012BE0F; EMBL:X17258; NID:g2760; PIDN:CAA35159.1; PID:g2761

A;Note: this is a revision to the sequence from reference S08240

R;de Oliveira Azevedo, M.; Radford, A.

Nucleic Acids Res. 18, 668, 1990

A;Title: Sequence of cbh-1 gene of *Humicola grisea* var. *thermoidea*.

A;Reference number: S08240; MUID:90175006; PMID:2308855

A;Accession: S08240

A;Molecule type: DNA

A;Residues: 1-299, 'H', 301-525 <DEO>

A;Cross-references: UNIPARC:UPI00001729F6; EMBL:X17258

A;Note: the authors translated the codon CAG for residue 87 as His

A;Note: this sequence has been revised in reference S38794

R;Azevedo, M.; de, O.; Felipe, M.S.S.; Astolfi-Filho, S.; Radford, A.

J. Gen. Microbiol. 136, 2569-2576, 1990

A;Title: Cloning, sequencing and homologies of the cbh-1 (exoglucanase) gene of *Humicola grisea* var. *thermoidea*.

A;Reference number: A45869; MUID:91178527; PMID:2127803

A;Accession: A45869
A;Status: not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-20, 'R', 22-34, 'K', 36-86, 'H', 88-141, 'V', 143-157, 'Y', 159-237, 'QQH', 241-244, 'I', 246-299, 'H', 301-525 <AZE>
A;Cross-references: UNIPARC:UPI00001729F7; GB:M64588; GB:X17258
A;Note: this sequence has been revised. See entry S08240
C;Genetics:
A;Gene: cbh-1
A;Introns: 138/1
C;Superfamily: cellulose 1,4-beta-celllobiosidase I; fungal cellulose-binding domain homology
C;Keywords: glycosidase; hydrolase; polysaccharide degradation
F;494-525/Domain: fungal cellulose-binding domain homology <FCB>

Query Match 59.9%; Score 1641.5; DB 1; Length 525;
Best Local Similarity 57.3%; Pred. No. 1.2e-91;
Matches 294; Conservative 74; Mismatches 122; Indels 23; Gaps 8;

Qy 61 CPD-NECAKNCCLDGAAYASTYGVTTSGNSLSIGFVTQSAQK-NVGARLYLMASDTTYQE 118
| | ||:|||:||| | ||||:|||:|||: |||: | ||| : | | | : | | |
Pb 78 CTDIAKSCAQNCCVHDGADYVTCYLTTCGDSLKLKEVYTKGQVSTVYICSPDTYLMPCGDVKYQT 138

Qy 119 FTLLGNEFSFDVDVSQLPCGLNGALYFVSMADGGVSKYPTNTAGAKYGTGYCDSQCPRD 178

Qy 179 LKFINGQANVEGWEPSNNANTGIGGHGSCCSEMDIWEANSISEALTPHPCTTVGQEICE 238

Qy 297 ---GAINRYYVQNGVTFQQPNAELGSYSGNELNDDYCTAEEAEFGG-SSFSDKGGLTQFK 352

D_b 317 GDLGEIKRFYQDGKIIIPNSESTIPEGVEGNSITQDWCDRQKVAFGDIDDFNRKGGMQMG 376
O_v 353 KATSCGMVILVMSLWDDYXANMILWLDSTYKRTNETSSTDGAIVPGCSCTSSCVRAQVESOSPN 412

Db 377 KALAGPMVLVMSIWDDHASNMLWLDSTFPV-DAAGKPGAERGACPTTSGVPAEVEAEAPN 435

Db 436 SNVVF SNIRFG PIGSTVAGLPGAGNGGNNGNPP--PPTTTSSAPATT TASAGPKAG 492

Qy 464 HYGQCGGIGYSGPTVCASGTTQVLPYYSQCL 496
 : |||||||::||| | | | || :|||||
Db 493 RWOOCGGIGFTGPTOCEEPYICTKLNDWYSQCL 525